

Inequality across US counties: an S factor analysis

Emil O. W. Kirkegaard*



Open Quantitative
Sociology & Political
Science

Abstract

A dataset of socioeconomic, demographic and geographic data for US counties (N3,100) was created by merging data from several sources. A suitable subset of 28 socioeconomic indicators was chosen for analysis. Factor analysis revealed a clear general socioeconomic factor (S factor) which was stable across extraction methods and different samples of indicators (absolute split-half sampling reliability = .85). Self-identified race/ethnicity (SIRE) population percentages were strongly, but non-linearly, related to cognitive ability and S. In general, the effect of White % and Asian % were positive, while those for Black %, Hispanic % and Amerindian % were negative. The effect was unclear for Other/mixed %. The best model consisted of White %, Black %, Asian % and Amerindian % and explained 41/43 % of the variance in cognitive ability/S among counties. SIRE homogeneity had a non-linear relationship to S, both with and without taking into account the effects of SIRE variables. Overall, the effect was slightly negative due to low S, high White % areas. Geospatial (latitude, longitude, and elevation) and climatological (temperature, precipitation) predictors were tested in models. In linear regression, they had little incremental validity. However, there was evidence of non-linear relationships. When models were fitted that allowed for non-linear effects of the environmental predictors, they were able to add a moderate amount of incremental validity. LASSO regression, however, suggested that much of this predictive validity was due to overfitting. Furthermore, it was difficult to make causal sense of the results. Spatial patterns in the data were examined using multiple methods, all of which indicated strong spatial autocorrelation for cognitive ability, S and SIRE (k nearest spatial neighbor regression [KNSNR] correlations of .62 to .89). Model residuals were also spatially autocorrelated, and for this reason the models were re-fit controlling for spatial autocorrelation using KNSNR-based residuals and spatial local regression. The results indicated that the effects of SIREs were not due to spatially autocorrelated confounds except possibly for Black % which was about 50 % weaker in the controlled analyses. Pseudo-multilevel analyses of both the factor structure of S and the SIRE predictive model showed results consistent with the main analyses. Specifically, the factor structure was similar across levels of analysis (states and counties) and within states. Furthermore, the SIRE predictors had similar betas when examined within each state compared to when analyzed across all states. It was tested whether the relationship between SIREs and S was mediated by cognitive ability. Several methods were used to examine this question and the results were mixed, but generally in line with a partial mediation model. Jensen's method (method of correlated vectors) was used to examine whether the observed relationship between cognitive ability and S scores was plausibly due to the latent S factor. This was strongly supported ($r = .91$, $N_{indicators}=28$). Similarly, it was examined whether the relationship between SIREs and S scores was plausibly due to the latent S factor. This did not appear to be the case.

Keywords: intelligence, IQ, cognitive ability, race, general socioeconomic factor, NAEP, S factor, United States, LASSO regression, inequality, precipitation, elevation, method of correlated vectors, spatial autocorrelation, SIRE, longitude, latitude, USA, temperature, spatial statistics, multilevel, Jensen's method

1 Introduction

Previous studies have examined the relationship between the composition of political units – in terms of self-identified race/ethnicity (SIRE) or other demographic variables – cognitive ability and socioeconomic outcomes (Beaver & Wright, 2011; Fuerst & Kirkegaard, 2016a,b; Kirkegaard, 2015c,a, 2016a; Kirkegaard & Fuerst, 2016; Pesta & Poznanski, 2014; Putterman & Weil, 2010). In general, it has been found that demographic

*University of Aarhus, Denmark, E-mail: emil@emilkirkegaard.dk

composition is a moderate to strong predictor of cognitive ability and socioeconomic outcomes. Perhaps the most comprehensive study is that of [Fuerst & Kirkegaard \(2016a\)](#) who estimated the genomic ancestry of every country in the Americas as well as the first-order administrative divisions¹ of the US, Mexico, Colombia and Brazil. European ancestry correlated .71 and .64 with estimated cognitive ability and the general socioeconomic factor (S factor), respectively. Studies that have examined administrative divisions within countries have usually, but not always, examined the first-order divisions. One exception is [Beaver & Wright \(2011\)](#) which examined counties in the US (second-order administrative divisions). Their study did not focus on the demographic correlations, but reported that African American % correlated with cognitive ability and S at -.43 and .77 (corrected for attenuation, .72 before).² They did not attempt to combine multiple demographic predictors nor did they examine geospatial or geographic predictors.

The purpose of this study is to build upon the earlier analyses by examining the structure of the S factor and its demographic and environmental correlates at the county-level in the US.

2 Data

To analyze inequality and correlates among US counties, several datasets were combined:

1. Measure of America's county data covering 2009-2010. <http://www.measureofamerica.org/download-agreement/>
2. Countyhealthrankings.org's health data for 2014. <http://www.countyhealthrankings.org/rankings/data>
3. Factfinder's geographic dataset for counties for 2010. http://factfinder.census.gov/bkmk/table/1.0/en/DEC/10_SF1/G001/0100000US.05000.003
4. Factfinder's age and gender dataset for counties for 2010. http://factfinder.census.gov/bkmk/table/1.0/en/ACS/10_5YR/S0101/0100000US.05000.003
5. NOAA 1980-2010 normals (climate data). Average annual data for the last 30 years. <http://www.ncdc.noaa.gov/cdo-web/datatools/normal>s
6. School-district data from George W. Bush Institute Global Report Card. <http://globalreportcard.org/about.html>

The dataset first contains the population proportions by SIRE as well as a number of socioeconomic variables. The second contains a large number of health-related variables as well as other socioeconomic variables. The third contains geographic information for the counties (latitude and longitude) which was needed for the spatial analysis and for the environmental models. The fourth contains very detailed population data, but was chosen because it included a median age variable which would otherwise have had to be estimated from age band variables.

The fifth dataset contains climatological data (precipitation, mean/min/max temperature) and other data (including geolocation) for weather stations in the US ($N \approx 8400$). These data are publicly available at the listed website, but not in a format that is easy to work with. To overcome this, I had to download the data automatically using a scraper script.³ Each station was then linked to the nearest county using distance calculated from the geolocation variables, and the data were averaged across stations for each county. This resulted in most counties having climatological data available ($N=2914$). This is almost the same method as that employed by [Deryugina & Hsiang \(2014\)](#).

The sixth dataset contains scholastic achievement (reading and math), student expenditure, SIRE composition, geolocation (latitude and longitude) and other data for school-districts. The achievement scores were given in

¹ These have different names, but are often called states, regions or the like.

² S was not extracted in the study. However, using matrix algebra it is possible to extract an S factor and calculate the correlation of African American % to this factor. See [Revelle \(2015a\)](#) for the method. Thanks to David Condon for help with this task. The correlation between cognitive ability and S was .64 (corrected for attenuation, .60 before). To extract the S factor, I used all the non-composite S variables (details can be found in the beaver.R file).

³ Details of the scraping can be found at <http://emilkirkegaard.dk/en/?p=5904>.

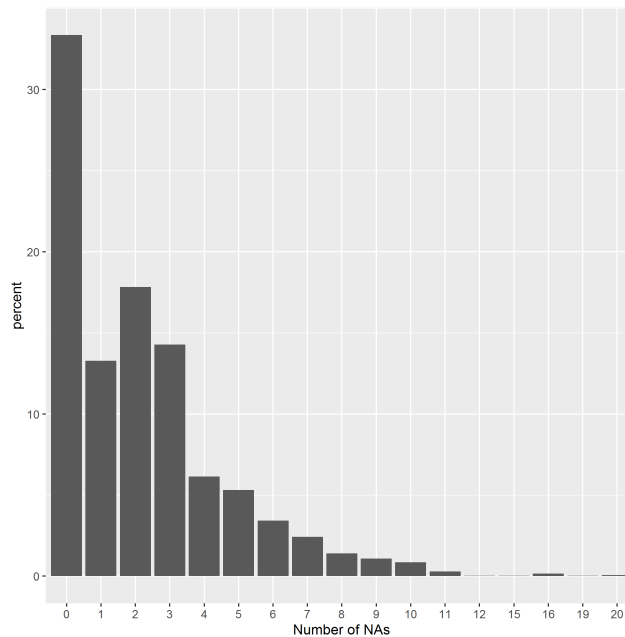


Figure 1: Histogram of missing data by case.

national centiles, which were converted to Z-scores. The dataset contains achievement scores for 2004-2009, but because the year-to-year correlations were near perfect (.99), only the 2009 data were used in this study.

Together, the datasets contain hundreds of variables, many of which are of a socioeconomic nature. To extract an S factor, it is necessary to choose a subset of these for analysis. I chose 31 variables based on three principles:

- 1) it must be an important socioeconomic outcome,
- 2) it must not be strongly dependent on local geography, and
- 3) the overall balance of variables should not be strongly skewed towards a particular domain.

Others would likely have chosen somewhat different variables. The effect of the choice of indicators is examined in Section 3.5.

Unless otherwise noted, all analyses were weighted using the total population in the county if technically possible.

3 S factor analysis

3.1 Missing values and imputation

For the purpose of extracting an S factor, there cannot be any missing data, so cases with missing data must either be excluded or have their missing data imputed. The distribution of missing data is shown in Figure 1.

Only .3 % of cases had more than 10 missing datapoints. One could perhaps impute the data for the cases with up to 20 missing datapoints, but these cases would have needed to have 67 % of their data imputed. Instead, I excluded cases with more than 10 missing datapoints. The missing data for the remaining cases (N=3127) were imputed using the `irmi` function from the **VIM** package for R with no noise component (Templ et al., 2015). The function uses robust methods so that the imputed values are sensible even in the case of outliers (Templ et al., 2011).

3.2 Redundancy analysis

Datasets of socioeconomic variables often contain variables that are perfectly related to each other, such as *proportion of high school graduates and proportion of non-high school graduates*. Including both variables can cause problems and in effect give double weight to one aspect of inequality. To counteract this, an algorithm was used that excludes variables until no pair of variables correlate (positively or negatively) above a specified cutoff (Kirkegaard, 2016b). The cutoff used in all previous S factor studies is .90 which was also used in this study. This resulted in 3 variables being excluded: 2 educational attainment variables out of 4 initial, and 1 poverty variable out of 4 initial.⁴

3.3 Extraction stability across methods

There are multiple methods to estimate factor loadings and to calculate factor scores from loadings. Sometimes, the results differ dramatically between these. If researchers do not examine whether the results are stable across factor extraction and scoring methods, they may be misled (Kirkegaard, 2016b). To guard against this, I extracted the first factor using all combinations of factor extraction and factor scoring methods available in the *fa* function in the **psych** package for R (Revelle, 2015b). After this, the factor scores were correlated to examine stability. The scores were very stable across methods with a mean correlation of 1.00.

3.4 Primary factor analyses

There are three aspects of extraction that were not already covered. I treat these as parameters and run the analyses for all possible parameter values ($2^3 = 8$ analyses in total).

First, one can extract the factor using weighted or unweighted analysis. When one uses weights, a matrix of weighted correlations is used as the basis of the analysis whereas normally one would use unweighted correlations.

Second, whether to use interval or rank-ordered data. Sometimes, variables have very non-normal distributions or non-linear relationships to other variables, both of which violate the assumptions behind correlations and hence factor analysis (the data must be multivariate normal; (Everitt & Hothorn, 2011)). One can counteract this to some degree by first converting the data to rank-ordered form. This loses some information because the distances between datapoints are set to be equal.

Third, whether to correct the variables for age. There are fairly large differences in median age between the counties. Both age and cohort (which are confounded in cross-sectional data) are known to be related to some variables such as educational attainment. One can attempt to control for this by regressing out the effect of age. Since I expected that the age ~ indicator relationships may be non-linear, I used local regression (LOESS, (James et al., 2013)), which is a kind of moving average function capable of capturing non-linear relationships. A model for each S indicator was fitted with the S indicator as the outcome variable and with median age as the predictor. The residuals of these models were then saved.

Figure 2 shows the factor loadings across the 8 analyses.

A clear S factor can be seen with no important indicator loading 'in the wrong direction' (Kirkegaard, 2014c). The results were very stable across parameters values (all factor congruence coefficients were 1.00). The violent crime variable only had a moderate negative loading (-.40) which is strange because violent crime is usually a strong correlate of other socioeconomic variables and cognitive ability (Beaver & Wright, 2011; Kirkegaard, 2014a,b, 2016a). I don't have a hypothesis for why this is the case.

3.5 Stability across choice of indicators

Most previous studies of both the general factor of cognitive ability (g) and S have found that they tend to be similar when extracted from different indicators. This is known as the indifference of the indicator (Jensen, 1998; Johnson et al., 2008; Kirkegaard, 2016b, 2014c; Major et al., 2011; Thorndike, 1987). As mentioned earlier, it is usually possible in an S factor analysis to pick other socioeconomic variables to use. For this reason, I devised a method for quantifying how large the effect of indicator sampling is (Kirkegaard, 2016b). The method consists simply of dividing the indicators into two sets at random, extracting the first factor in each set

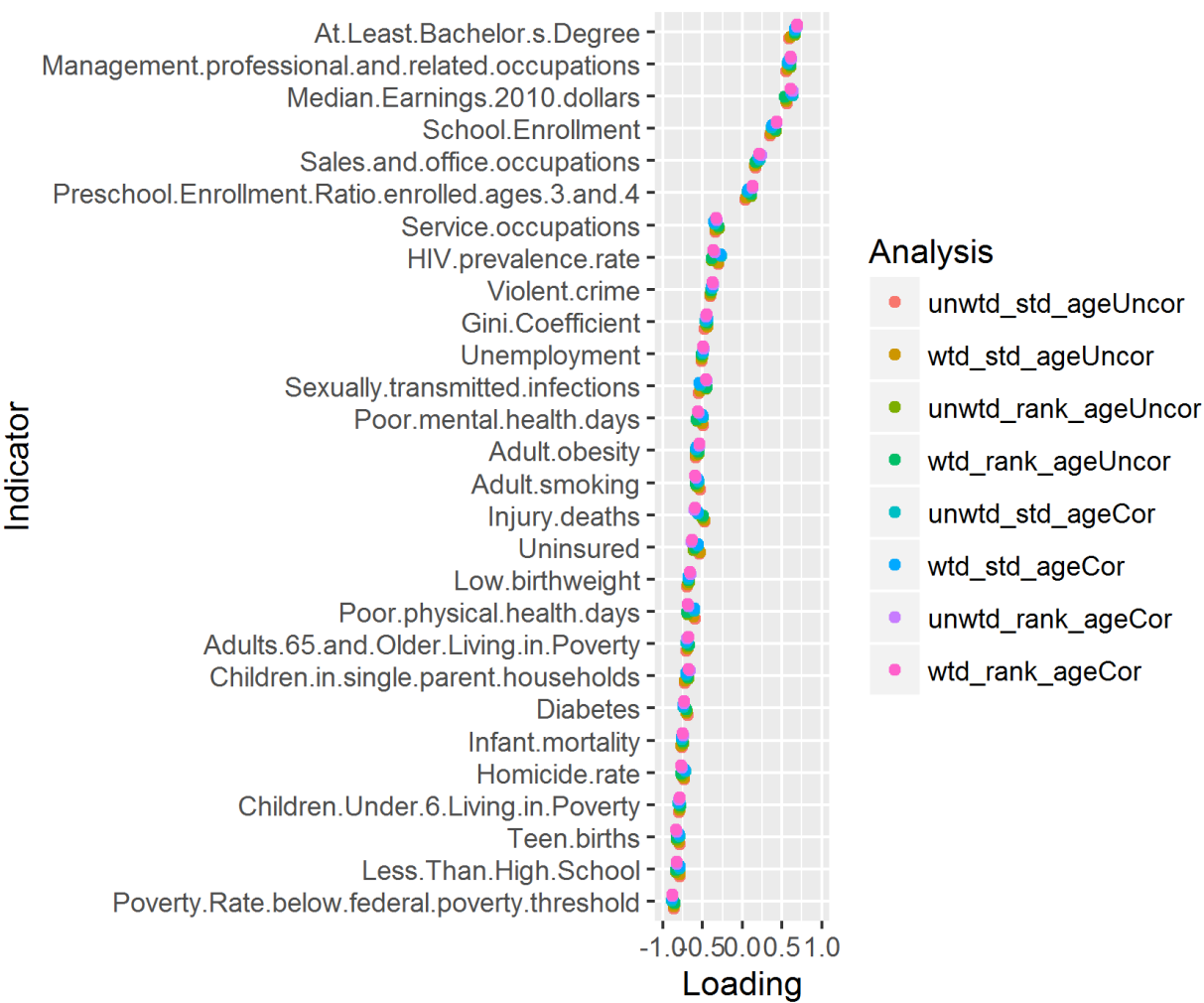


Figure 2: S factor loadings across parameter values.

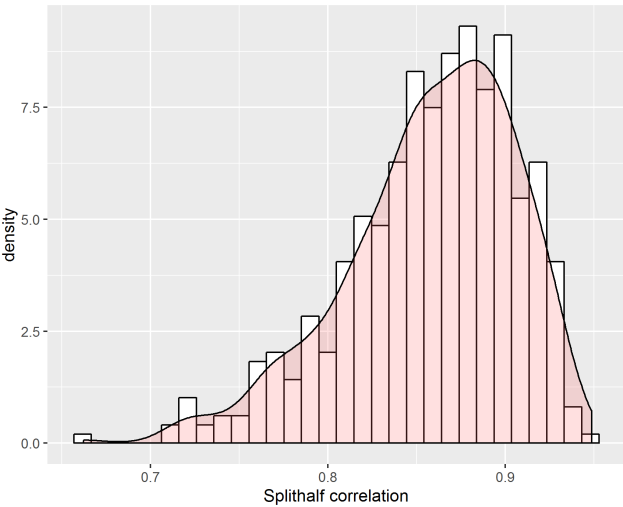


Figure 3: Distribution of split-half correlations. N=500.

independently, and correlating the scores. This process is then repeated a large number of times to obtain a stable result. Figure 3 shows the distribution of these correlations for the current dataset.

We see that the scores were generally very similar when they were extracted from different sets of indicators. The mean correlation was .86.

4 Self-identified race/ethnicity as predictors of CA/S

Previous studies of the relationship between SIRE and S in the US at the state-level found somewhat conflicting results. While there were usually moderate to strong negative relationships between Black % (percent of the population that is Black/African American; I will leave out the % for convenience) and S indicators and moderate to strong positive relationships between White/European and S indicators, in some but not all controlled analyses, the relationship was substantially reduced, disappeared entirely or even reversed in direction (Fuerst & Kirkegaard, 2016a,b; Pesta, 2016). The conflicting results may be due to aggregation effects (lack of ergodicity), that is, analyzing the data at too high a level. If there is an aggregation effect at the state-level that causes the conflicting results for White and S, then it will not be present when the data is analyzed at the county-level since this is one level below the state-level.

In contrast, the results for CA and White have generally been consistently positive, including when control variables are included in regressions.

Finally, the previous inconsistent results may simply be due to sampling error (Hunter & Schmidt, 2004). There are only 50 states in the US, so this is a small sample that results in considerable imprecision in the estimates.

Figures 4 and 5 show the relationships between SIREs (from the 2010 census⁶) and CA/S using LOESS. I only plot the fitted lines because plotting all the points would be chaotic (the plots with the points can be found in the supplementary materials).

There are clear non-linearities for some SIREs. In general, however, we see that White and Asian is associated with higher CA/S up to a point. The dip for White at the end is not a coincidence. The counties that are nearly 100 % White are mainly rural areas that span a wide range of CA/S levels. As expected, Black, Hispanic and Amerindian are clearly associated with lower CA/S and only show small deviations from linearity at the low percents. The odd patterns for Asian and Other are hard to make sense of and the wide error bars indicate that they may be a coincidence.

Since correlations are measures of linear associations, they can be misleading when the relations are non-linear. However, because most previous research on the topic has used correlations, the correlations are given in Table 1 for comparison.

Table 1: Intercorrelations between SIRE % variables and CA/S. Weighted correlation above the diagonal, unweighted below.

	CA	S	White	Black	Hispanic	Asian	Amerindian	Other
CA		0.68	0.59	-0.40	-0.40	-0.10	-0.13	-0.11
S	0.71		0.34	-0.52	-0.15	0.33	-0.14	0.15
White	0.57	0.53		-0.47	-0.76	-0.49	-0.05	-0.18
Black	-0.53	-0.57	-0.62		-0.10	-0.05	-0.09	-0.10
Hispanic	-0.14	-0.11	-0.58	-0.11		0.30	-0.03	0.00
Asian	0.06	0.23	-0.26	0.01	0.14		-0.08	0.54
Amerindian	-0.23	-0.20	-0.30	-0.10	-0.04	-0.01		0.10
Other	-0.08	0.07	-0.17	-0.09	-0.03	0.42	0.26	

In general, we see the expected patterns. White is a notably weaker predictor of S when using weighted correlations (.53 vs. .34). This is related to the dip at the end of the line seen in Figure 5. The relationship

⁴ The excluded variables were: At.Least.High.School.Diploma (reverse coded version of Less.Than.High.School, $r=-.993$), Child.Poverty.living.in.families.below.the.poverty.line (redundant with Poverty.Rate.below.federal.poverty.threshold, $r=.924$) and Graduate.Degree (redundant with At.Least.Bachelor.s.Degree, $r=.922$).

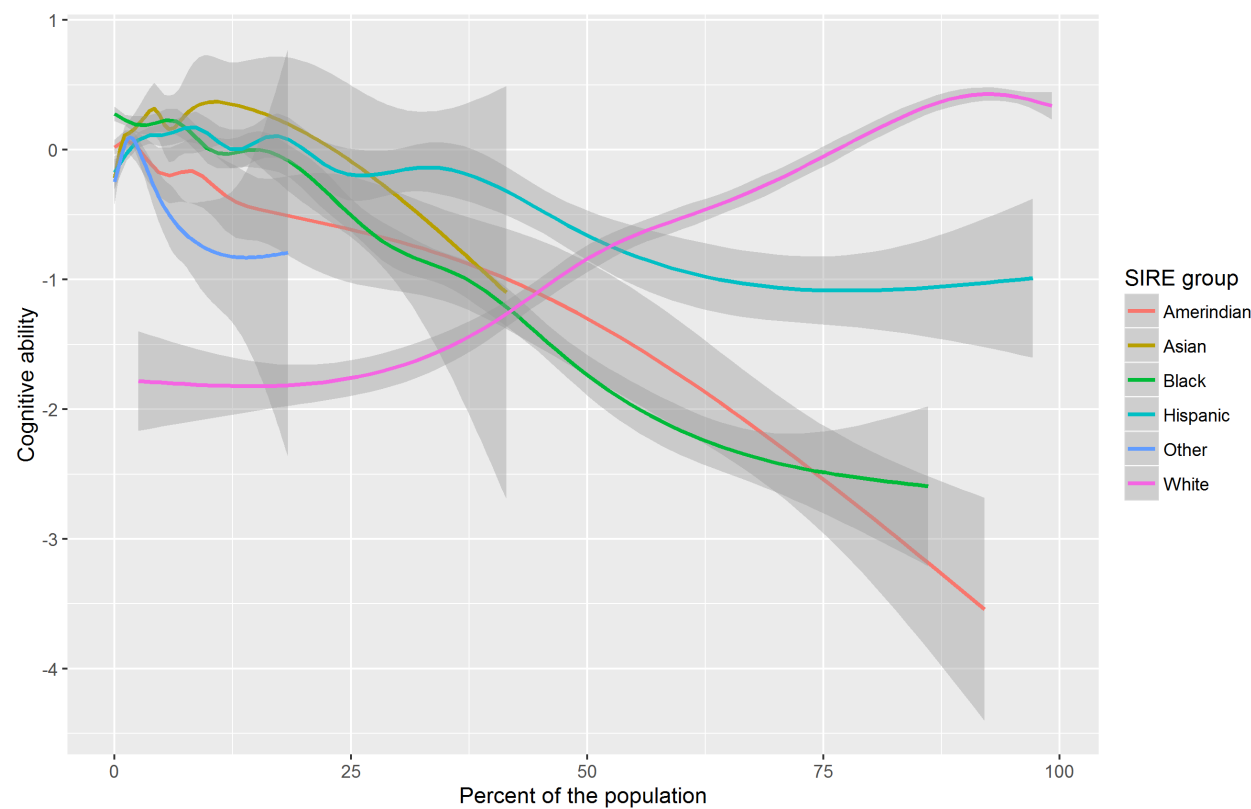


Figure 4: Relationships between SIRE % of counties and cognitive ability. Error bars are 95 % analytic confidence intervals. Lines are from LOESS regressions.

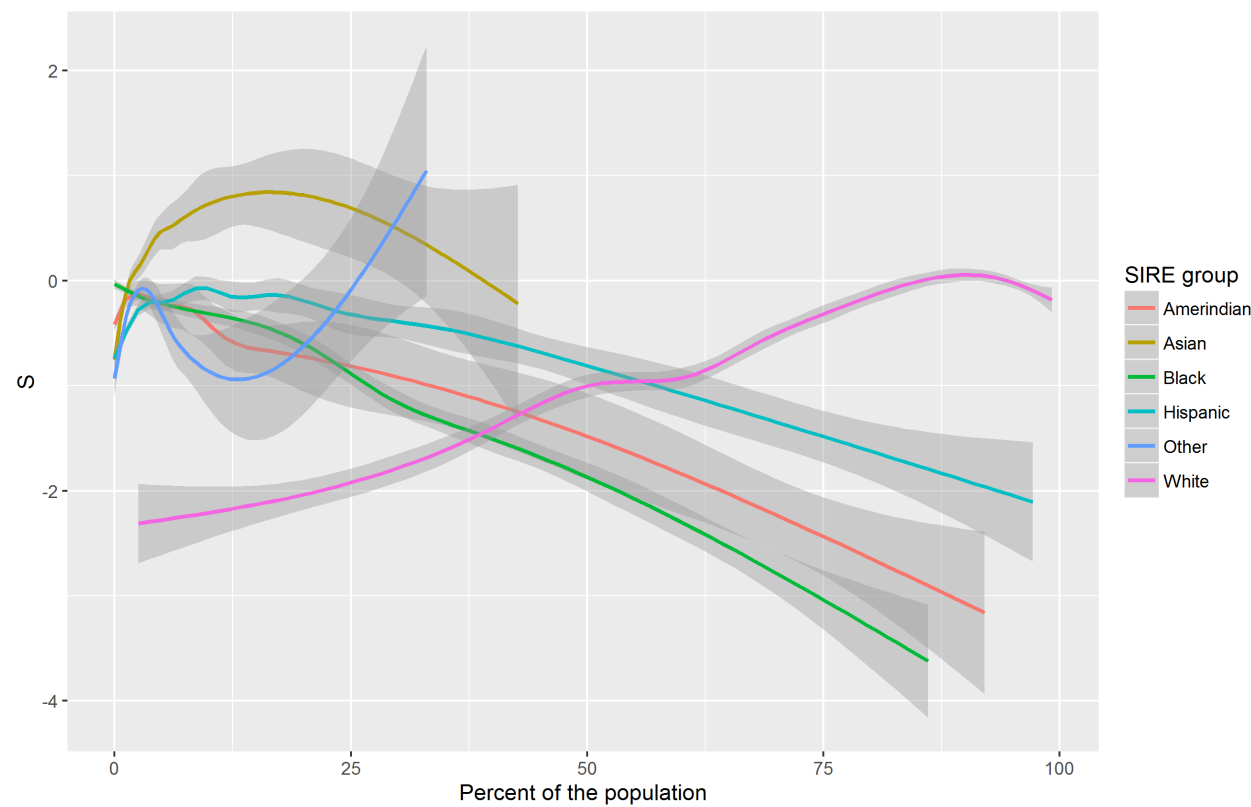


Figure 5: Relationships between SIRE % of counties and S scores. Error bars are 95 % analytic confidence intervals. Lines are from LOESS regressions.

between Black and S was weaker in the present study than it was in [Beaver & Wright \(2011\)](#)'s (-.72 uncorrected). Still, the only really odd result is the weak negative correlation between Asian and CA when using weights (-.10).

4.1 Multiple linear regression with SIREs

Since the SIRE variables must sum to 100 %, taken together, the variables are perfectly linearly dependent and thus cannot all be included in a regression model. However, it is possible to include some of them and leave others out. As was done in prior analyses ([Fuerst & Kirkegaard, 2016a](#); [Kirkegaard & Fuerst, 2016](#)), I tested all the possible models (63 models; $N_{models} = p^2 - 1$, where p is the number of predictors). The top 5 models according to BIC are shown in Tables 2 and 3. All the models are shown in the supplementary materials (sheet CA_SIRE_models and S_SIRE_models).

Table 2: Top 5 SIRE models according to BIC. Outcome: CA.

White	Black	Hispanic	Asian	Amerindian	Other	BIC	R adj.
0.72		0.11	0.14		-0.24	10653.79	0.64
-0.38	-0.81	-0.62		-0.47	-0.33	10657.27	0.64
-14.48	-11.15	-9.92	-1.75	-5.80	-1.70	10657.60	0.64
	-0.53	-0.37	0.05	-0.33	-0.29	10657.70	0.64
0.55	-0.12		0.12	-0.12	-0.24	10658.33	0.64

Table 3: Top 5 SIRE models according to BIC. Outcome: S.

White	Black	Hispanic	Asian	Amerindian	Other	BIC	R adj.
0.70		0.20	0.26		-0.06	10191.82	0.66
1.26	0.41	0.57	0.33	0.19		10199.44	0.66
0.74	0.03	0.23	0.27		-0.05	10199.53	0.66
0.71		0.21	0.25			10199.54	0.66
0.70		0.20	0.26	-0.02	-0.05	10199.54	0.66

One problem with selecting the top models this way is that one sometimes selects nonsensical models like models #2-3 for predicting CA. The strange betas are related to the lack of free variation in the predictors (multicollinearity). A better way to summarize the findings is to summarize the betas across all models, as done in Tables 4 and 5.

Table 4: Summary statistics of predictor betas across all models. Dependent variable: CA. Predictors: SIREs.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other
mean	0.13	-0.63	-0.42	0.01	-0.41	-0.23
median	0.56	-0.46	-0.31	0.06	-0.28	-0.21
sd	2.71	1.96	1.78	0.33	1.01	0.28
mad	0.14	0.08	0.07	0.11	0.13	0.08

Table 5: Summary statistics of predictor betas across all models. Dependent variable: S. Predictors: SIREs.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other
mean	0.41	-0.44	-0.03	0.22	-0.25	0.02
median	0.41	-0.53	-0.13	0.20	-0.29	-0.03
sd	0.77	0.55	0.51	0.10	0.29	0.13
mad	0.31	0.06	0.32	0.10	0.12	0.16

Because some of the models are nonsensical, it is best to use robust statistics (medians, median absolute deviations [mad]). Looking at these we see much the same results as we saw in Figures 4 and 5: White and Asian (only slightly for CA) are associated with better outcomes, Black, Hispanic, Amerindian and Other are associated with worse outcomes.

In contrast to earlier findings (Fuerst & Kirkegaard, 2016a; Kirkegaard & Fuerst, 2016), using more than one SIRE predictor did improve the predictive validity for S. The best single-predictor model for S is Black % which has an adjusted R of .52, but the best multi-predictor models had .66. For cognitive ability, the best single-predictor model was White with an adjusted R of .59 while the best model had .64 which is only a slight improvement.

Trying all the models and picking the best by some model fit indicator tends to result in overfitted models (James et al., 2013). LASSO regression gets around this problem by systematically shrinking parameters towards 0. The amount of shrinkage is determined using in-sample cross-validation (k-fold). This results in simpler models that are more likely to retain their predictive validity in a new sample. LASSO regression with cross-validation, however, has a random component (the folds are random), so results are not identical across runs. I ran it 500 times and summarized the results, which are shown in Tables 6 and 7.

Table 6: Summary statistics of predictor betas from 500 LASSO regression runs. Dependent variable: CA. Predictors: SIREs.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other
mean	0.42	-0.10	0.00	0.02	-0.06	0.00
median	0.42	-0.10	0.00	0.02	-0.06	0.00
sd	0.02	0.01	0.00	0.01	0.02	0.01
mad	0.02	0.00	0.00	0.01	0.02	0.00
fraction zero/NA	0.00	0.00	1.00	0.06	0.04	0.58

Table 7: Summary statistics of predictor betas from 500 LASSO regression runs. Dependent variable: S. Predictors: SIREs.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other
mean	0.25	-0.28	0.00	0.15	-0.06	0.00
median	0.26	-0.28	0.00	0.15	-0.07	0.00
sd	0.02	0.00	0.00	0.01	0.02	0.00
mad	0.02	0.00	0.00	0.01	0.02	0.00
fraction zero/NA	0.00	0.00	1.00	0.00	0.00	0.99

LASSO regression generally confirmed the findings from before, namely that White and Asian were positive predictors, while Black and Amerindian were negative predictors. It found no effect of Hispanic and Other, similar to their weak and null values before. I will hence refer to the model with White, Asian, Black and Amerindian as the best (SIRE) model.

4.2 SIRE homogeneity

Previous research has found very mixed results with regards to the effects of diversity on socioeconomic measures (e.g. interpersonal trust) (Eagly, 2016; Meer & Tolsma, 2014). It is possible to construct a measure of SIRE homogeneity (the opposite of heterogeneity/diversity) with the data using what is often called Simpson's index. This is the chance that two randomly picked cases will belong to the same group, and is a measure of variation for nominal data that is bounded between 0 and 1 (least and most homogeneous with regards to SIRE). Figure 6 shows a LOESS plot of the relationship between homogeneity and S.

SIRE homogeneity has a non-linear relationship to S. The lowest S is found at both mid levels of homogeneity (around .5) and very high levels (close to 1). The latter part is presumably connected to the dip of White in Figure 5.

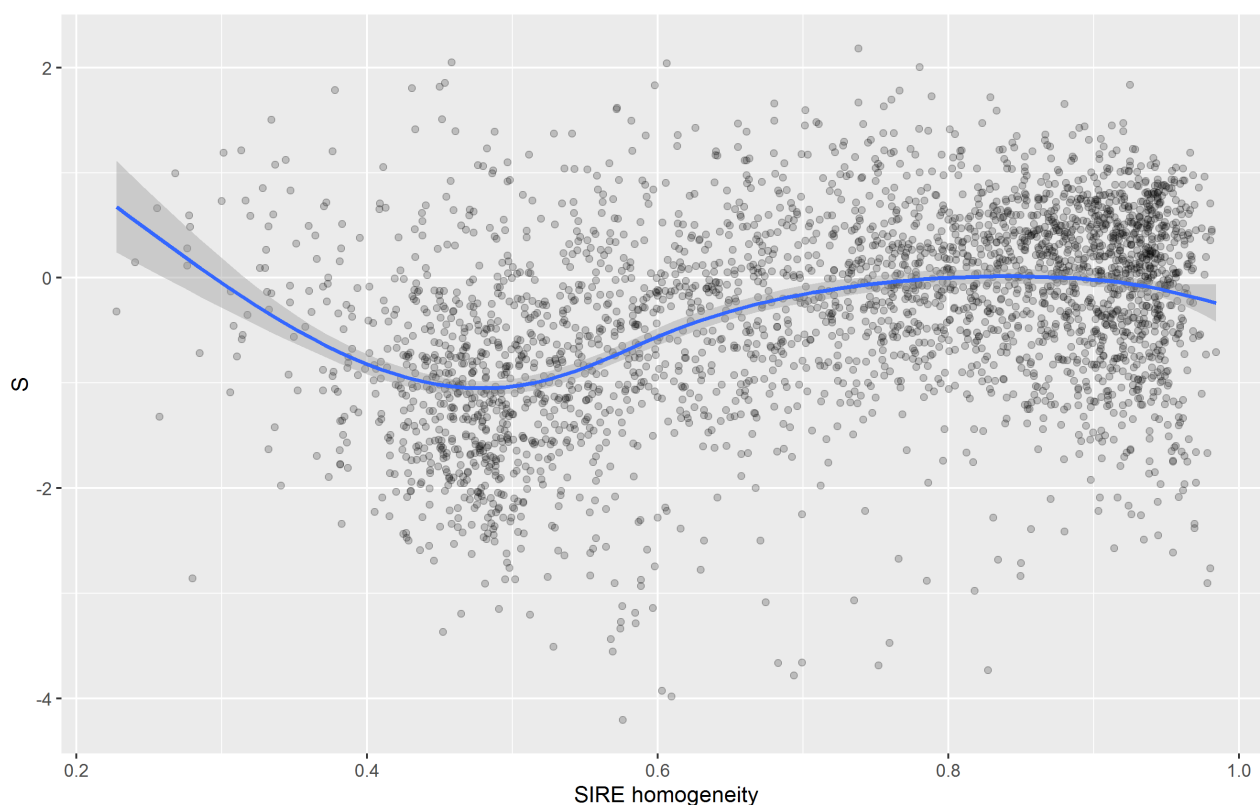


Figure 6: The relationship between SIRE homogeneity and S. Line fitted by LOESS.

The odd pattern may be due to confounding with SIRE variables. To test this, the same approach as above was used except that the SIRE homogeneity variable was included among the predictors. Tables 8 and 9 show the summary statistics for the models. The full list of models can be found in the supplementary materials (sheet S_SIRE_homogeneity_models).

Table 8: Summary statistics for all models. Dependent variable: S. Predictors: SIREs + SIRE homogeneity.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other	SIRE homogeneity
mean	0.55	-0.49	-0.03	0.22	-0.25	0.00	-0.16
median	0.55	-0.53	-0.09	0.23	-0.28	-0.05	-0.29
sd	0.68	0.49	0.47	0.08	0.26	0.13	0.31
mad	0.36	0.18	0.42	0.07	0.16	0.15	0.28

Table 9: Summary statistics of predictor betas from 500 LASSO regression runs. Dependent variable: S. Predictors: SIREs + SIRE homogeneity.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other	SIRE homogeneity
mean	0.34	-0.30	0.00	0.17	-0.13	-0.01	-0.05
median	0.34	-0.30	0.00	0.17	-0.13	-0.01	-0.05
sd	0.04	0.01	0.00	0.01	0.02	0.01	0.04
mad	0.04	0.01	0.00	0.00	0.01	0.01	0.05
fraction zero / NA	0.00	0.00	1.00	0.00	0.00	0.18	0.18

The SIRE results are similar to the previous results and the results for SIRE homogeneity depended on which method was used: they were moderately negative using all OLS and only barely negative with LASSO (and was a redundant predictor in 18 % of the runs).

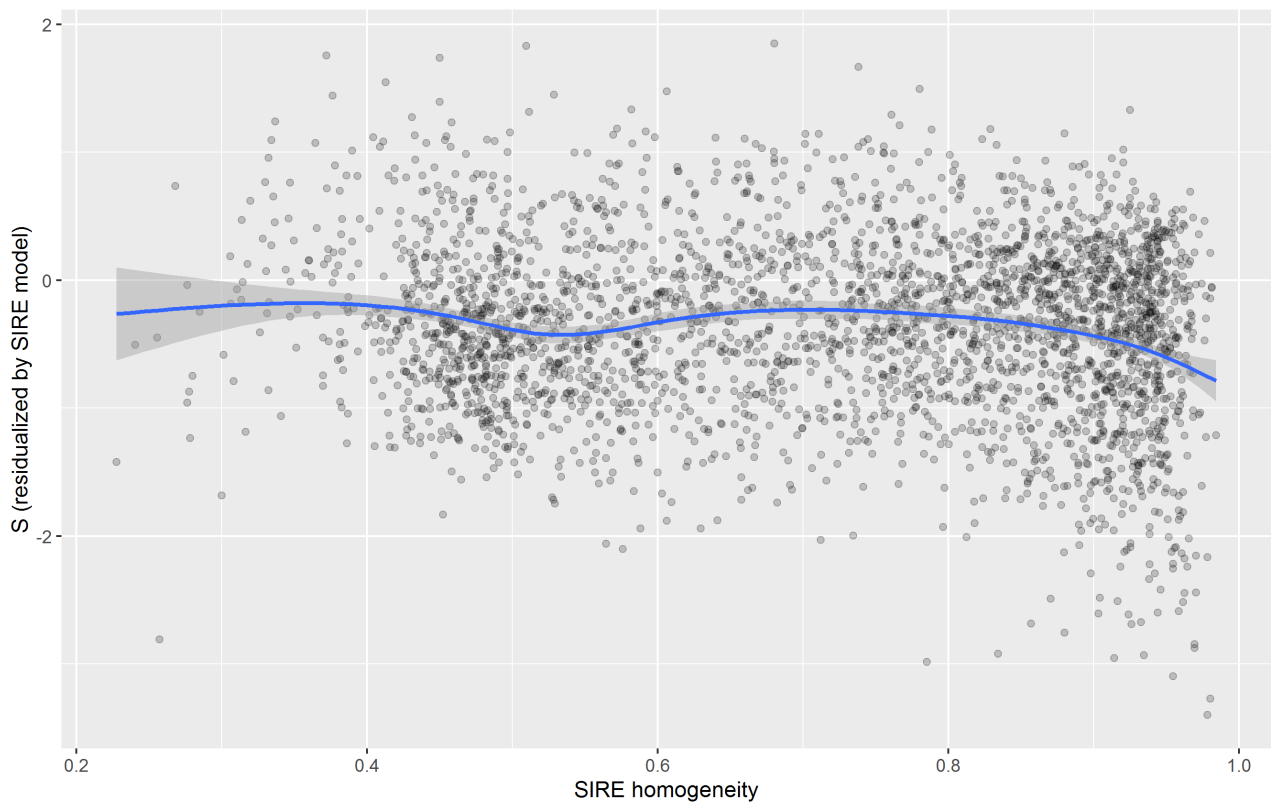


Figure 7: SIRE homogeneity and residualized S scores. Fitted line = LOESS.

It is possible that SIRE homogeneity has non-linear effects that persist after taking SIREs into account. To investigate this, I used the best SIRE model for S and extracted the residuals. Figure 7 shows the relationship between the residualized S scores and SIRE homogeneity.

There was a somewhat non-linear and slight negative relationship, as seen before the residualization. It is difficult to come up with a plausible causal model for the pattern and for that reason I did not include this as a predictor in later models.

4.3 Structural outliers and SIRE homogeneity

Exploratory factor analysis can reveal the underlying structure of a dataset (Everitt & Hothorn, 2011). However, some cases may not fit the structure well. These cases can be said to be structural outliers or mixed (Kirkegaard, 2016b). Previously, three main methods were developed to quantify this:

- 1) mean absolute residual (MAR),
- 2) change in factor size (CFS), and
- 3) absolute loading change (ALC) (Kirkegaard, 2016b).

Two variants of the second and third methods exist and each of the five methods-variants were applied to the present dataset. There was moderate to strong agreement between the methods (r 's .48 to .80). A general estimate was calculated by summing the Z-transformed scores on MAR, absolute CFS and mean ALC.

Mixedness in S factor analyses in general probably arises from aggregation effects where very different groups of persons are found in the same unit; when one takes the average of this, one will get strange numbers. For instance, the most mixed case in the present dataset is the District of Columbia, which houses both the very upper class group of persons in charge of the US and the lobbyists they attract, as well as a large Black underclass consisting of about 49 % of the population.

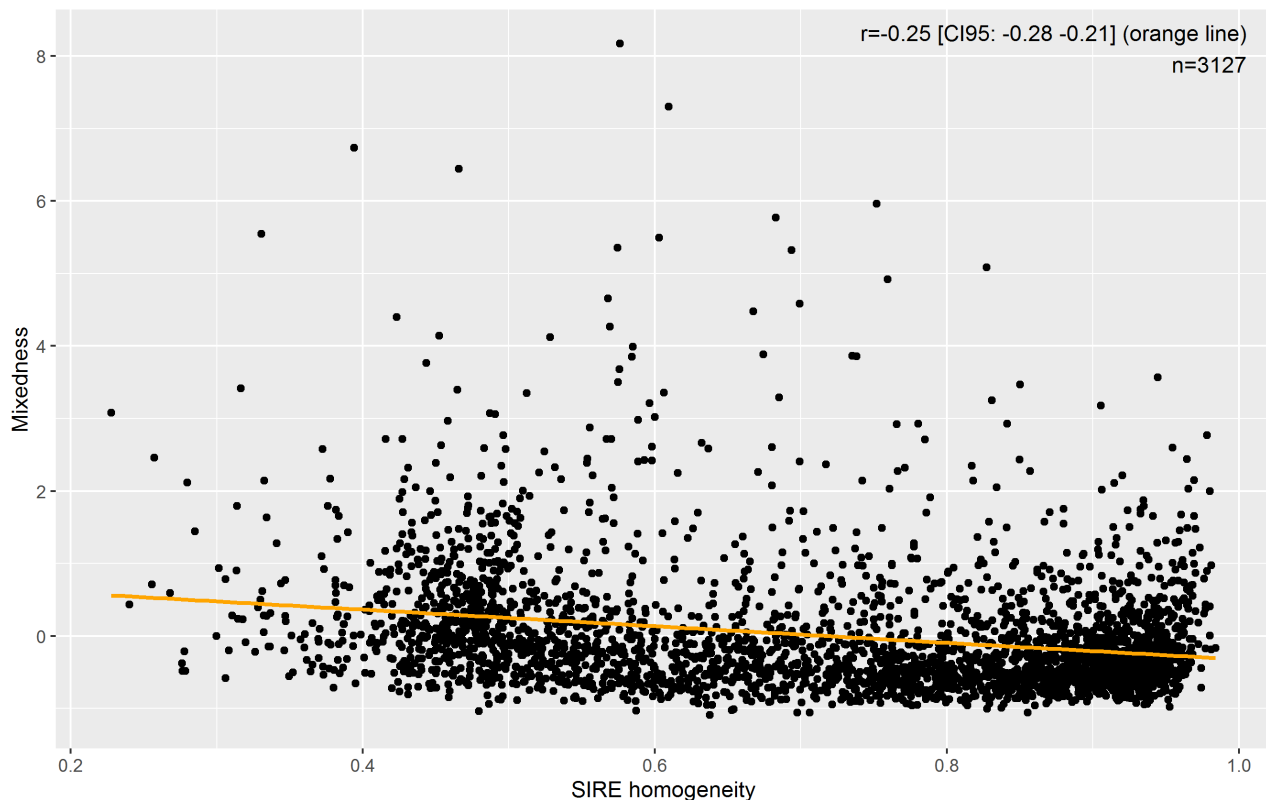


Figure 8: The relationship between SIRE homogeneity and S. Line fitted by LOESS.

It seems likely that such strange mixes of people will be found in counties that are more SIRE heterogeneous. Figure 8 shows the correlation of the general estimate of mixedness and SIRE homogeneity.

We see a relationship in the expected direction. It is also clear that a few cases are distant outliers with regards to mixedness. They may be interesting to examine in detail in a future study.

5 Natural environmental predictors of CA/S

Some have argued that the contemporary climate and geography have causal effects on cognitive ability and socioeconomic outcomes (León, 2015; León & Burga-León, 2015; León, 2016; Deryugina & Hsiang, 2014). Proponents of such theories have analyzed national-level or US state-level datasets. The study by Deryugina & Hsiang (2014) is an especially convincing exception. They used a multi-year dataset of daily temperatures for US counties. They looked at the years where a given county had many warm or cold days relative to the other years. It is thus a kind of repeated measurements model with random treatment/intervention assignments (year to year variation in temperature). The method thus controls for all fixed effects of the counties such as ethnoracial composition and geospatial location. If temperature has a causal effect on income, the counties that happen to have higher temperatures a given than normally a given year should also see a change in its income, and such a relationship was in fact found. The relationship was non-linear in the expected way: both very warm and very cold days had a negative impact on income. Furthermore, the result held across a number of controls such as taking into account the effect of temperature on neighboring counties. Most convincingly, they found that temperature had large effects on weekdays and small, even reversed effects on weekends.

I do not have daily or yearly temperature data, so the method used by Deryugina & Hsiang (2014) cannot be used on the present dataset. However, I can use the more traditional and methodologically weaker regressions methods employed before.

5.1 Linear regressions and correlations

The studies by León and my own previous studies all modeled the effects of climatological variables as being linear (Fuerst & Kirkegaard, 2016a,b; Kirkegaard & Fuerst, 2016; León, 2015; León, 2016; León & Burga-León,

2015). This is probably a bad idea given the results obtained by [Deryugina & Hsiang \(2014\)](#). However, since most previous studies used linear models, it is worth also doing the linear regressions and correlations for comparison purposes. Table 10 shows the intercorrelations between geospatial, climatological and outcome variables.

Table 10: Intercorrelations between outcome and climatological variables. Weighted correlations. Temperature data are annual values.

	CA	S	lat	lon	elevation	precip	temp	temp_max	temp_min
CA	1.00	0.68	0.32	0.38	-0.07	0.25	-0.37	-0.40	-0.32
S	0.68	1.00	0.34	-0.05	0.08	-0.10	-0.34	-0.34	-0.32
lat	0.32	0.34	1.00	0.10	0.12	-0.08	-0.93	-0.93	-0.90
lon	0.38	-0.05	0.10	1.00	-0.41	0.56	-0.20	-0.27	-0.12
elevation	-0.07	0.08	0.12	-0.41	1.00	-0.56	-0.24	-0.11	-0.36
precip	0.25	-0.10	-0.08	0.56	-0.56	1.00	0.05	-0.06	0.17
temp	-0.37	-0.34	-0.93	-0.20	-0.24	0.05	1.00	0.98	0.98
temp_max	-0.40	-0.34	-0.93	-0.27	-0.11	-0.06	0.98	1.00	0.92
temp_min	-0.32	-0.32	-0.90	-0.12	-0.36	0.17	0.98	0.92	1.00

The temperature variables (temp, temp_max, temp_min) were too strongly correlated to be used together in regression analysis. I used only the average temperature (temp) in the later analyses.

Tables 11 and 12 shows the OLS regression results for the best SIRE model with the climatological and geospatial variables added.

Table 11: Regression results. Outcome: CA. Predictors: SIREs, climatological and geospatial variables. CI = 95 % analytic confidence intervals. Cross-validated R = .72.*

Predictor	Beta	CI lower	CI upper
White	0.31	0.27	0.35
Black	-0.42	-0.46	-0.38
Asian	0.10	0.08	0.11
Amerindian	-0.14	-0.20	-0.09
lat	0.31	0.20	0.41
lon	0.40	0.37	0.44
elevation	0.17	0.12	0.22
precip	0.16	0.12	0.20
temp	0.27	0.16	0.38

*Cross-validated R's were calculated from 10-fold cross validation with 100 replications. See MOD_summary in the **kirkegaard** package for details.

Table 12: Regression results. Outcome: S. Predictors: SIREs, climatological and geospatial variables. CI = 95 % analytic confidence intervals. Cross-validated R = .71.

Predictor	Beta	CI lower	CI upper
White	0.29	0.25	0.33
Black	-0.37	-0.41	-0.33
Asian	0.23	0.22	0.25
Amerindian	-0.22	-0.28	-0.17
lat	0.35	0.27	0.43
lon	0.20	0.17	0.23
elevation	0.19	0.14	0.24
precip	0.02	-0.01	0.05
temp	0.28	0.19	0.36

We see that the betas for the SIRE predictors were mostly unchanged. Four of the new predictors seemed to have reliable predictive validity with betas from .19 to .35. Only precipitation did not seem to have any validity. However, the incremental validity added by the new predictors was negligible to small: delta R's .12 and .03, for cognitive ability and S, respectively.

Because OLS tends to overfit, the same predictors were examined using LASSO regression (500 runs). Table 13 and 14 shows the summary statistics.

Table 13: Summary statistics for LASSO regression. 500 runs. Outcome: cognitive ability. Predictors: SIREs, climatological and geospatial predictors.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other	lat	lon	elevation	precip	temp
mean	0.29	-0.33	0.00	0.05	-0.05	0.00	0.03	0.28	0.00	0.07	0.00
median	0.29	-0.33	0.00	0.05	-0.05	0.00	0.03	0.28	0.00	0.06	0.00
sd	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.00
mad	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.00
fraction zero/NA	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00

Table 14: Summary statistics for LASSO regression. 500 runs. Outcome: S scores. Predictors: SIREs, climatological and geospatial predictors.

Statistic	White	Black	Hispanic	Asian	Amerindian	Other	lat	lon	elevation	precip	temp
mean	0.22	-0.30	0.00	0.15	-0.11	0.00	0.09	0.02	0.00	0.00	0.00
median	0.22	-0.30	0.00	0.15	-0.11	0.00	0.09	0.02	0.00	0.00	0.00
sd	0.01	0.01	0.00	0.01	0.02	0.00	0.01	0.01	0.00	0.00	0.00
mad	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.02	0.00	0.00	0.00
fraction zero/NA	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.13	1.00	1.00	0.44

We see that LASSO regression agreed that the SIRE predictors were useful and that precipitation was useless, but that it disagreed with OLS with regards to the status of elevation which OLS found to be useful while LASSO did not. The status of temperature was very doubtful in LASSO regression, being redundant in 100/44 % of the runs for cognitive ability/S, and having a very small beta when it was not. The geospatial variables were also found to have small to near-zero betas in LASSO regression.

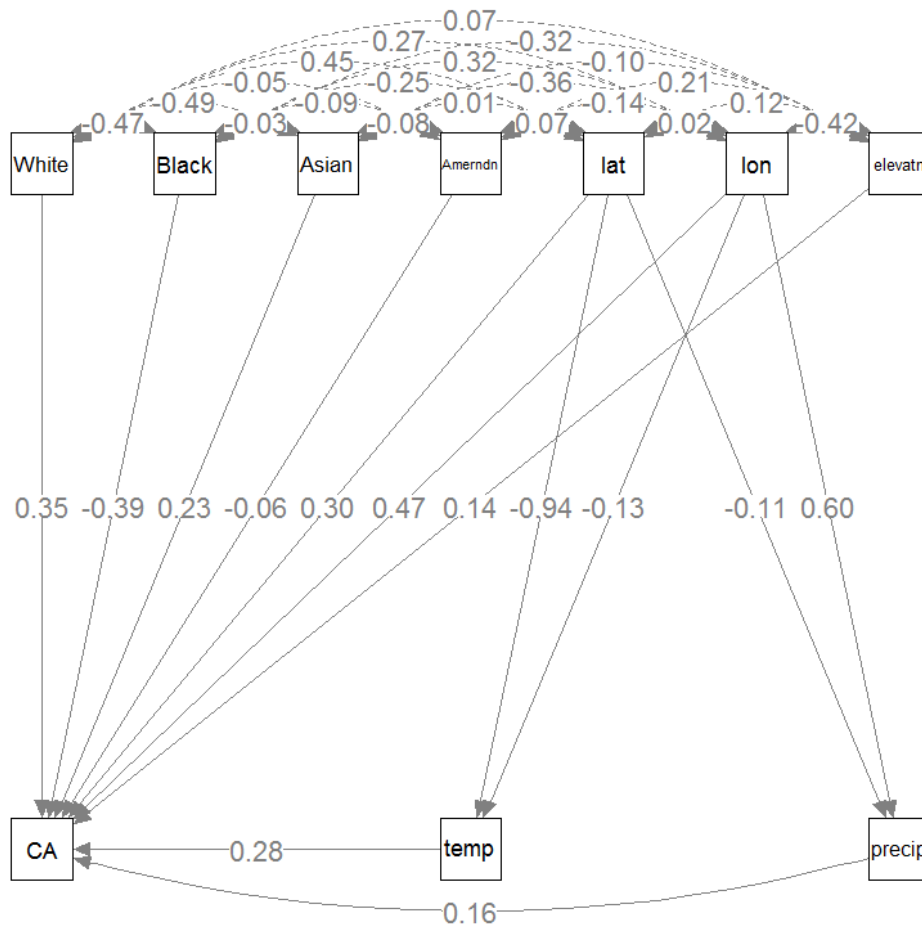


Figure 9: Path model for SIRE, climatological and geospatial variables' effect on cognitive ability.

5.1.1 Path modeling

The geospatial variables are causally antecedent to the climatological variables (mean temperature cannot cause longitude). Thus, it would make more sense to model the data in a way that takes into account this relationship between the predictors. This was done via path analysis fitted with the **lavaan** package (Rosseel et al., 2015).⁵ Figures 9 and 10 shows the path model with the estimated paths.

We see that the effects of the SIREs are similar to before with Asian and White both being stronger than what OLS and LASSO regression found. The four environmental variables also have positive paths as before and precipitation had around 0.

5.2 Non-linear effects of temperature

How non-linear are the effects of temperature in the present dataset? Figures 11, 12, 13 and 14 show the LOESS fitted lines for the relationship between annual mean temperature, and the outcome variables and the outcome variables residualized by the best SIRE model, respectively.

We see that temperature has a strong non-linear relationship to the outcome variables, both when they have been residualized by the SIRE model first and when they have not. Thus, the influence of temperature should be modeled using a model that allows for non-linear relationships. The method used by Deryugina & Hsiang (2014) was to discretize the temperature data and treat each bin as a level in a categorical variable. This allows for the non-linear effects to be seen. The same method was used in the present study. 10 equal-sized bins were

⁵ I additionally used the **survey** and **lavaan.survey** packages to allow for weighted path analysis (Lumley, 2014; Oberski, 2015).

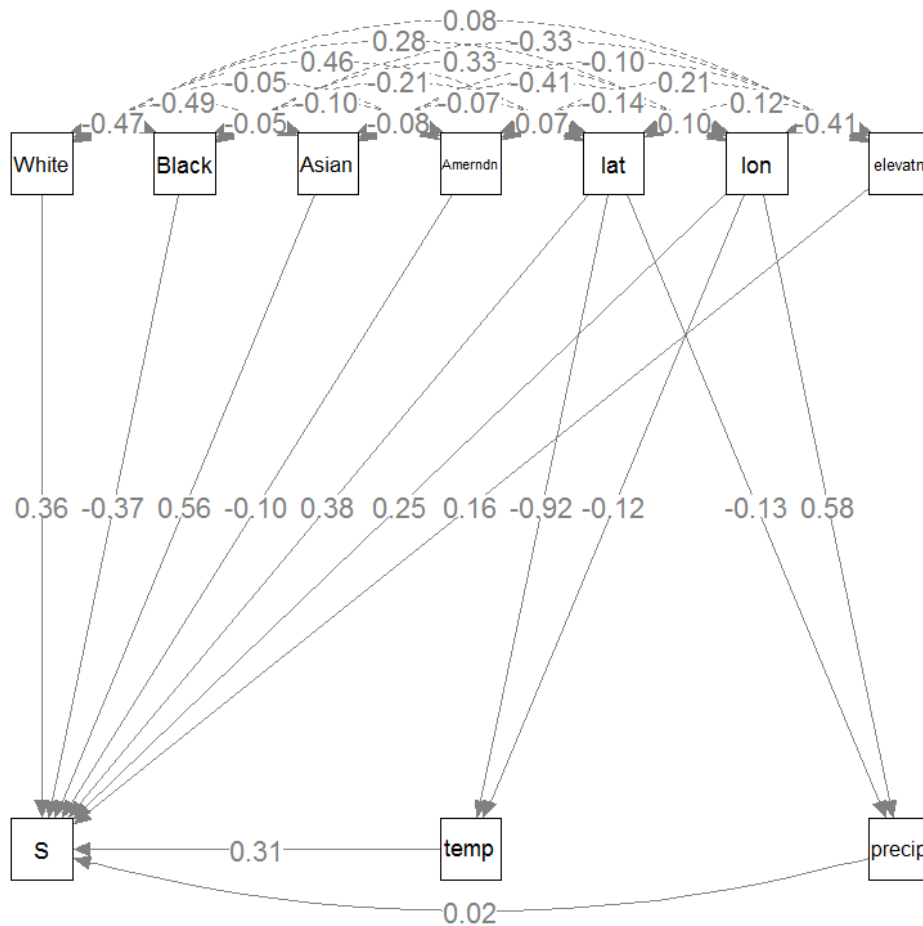


Figure 10: Path model for SIRE, climatological and geospatial variables' effect on S.

created. This was used instead of equal-ranged bins (the default) because the data were skewed, and using equal-ranged bins resulted in some bins having very few (<10) cases. Tables 15 and 16 shows the model betas.

Table 15: OLS regression betas. Outcome: cognitive ability. Predictors: SIREs and discretized temperature. Cross-validated $R = .65$.

Predictor	Beta	CI lower	CI upper
White	0.49	0.45	0.53
Black	-0.09	-0.13	-0.06
Asian	0.09	0.08	0.10
Amerindian	-0.18	-0.24	-0.11
temp_bins: (6.53,8.17]	0.02	-0.18	0.23
temp_bins: (8.17,9.56]	0.03	-0.16	0.22
temp_bins: (9.56,10.7]	0.06	-0.13	0.24
temp_bins: (10.7,11.9]	0.32	0.13	0.51
temp_bins: (11.9,13.4]	-0.01	-0.20	0.19
temp_bins: (13.4,14.9]	-0.39	-0.58	-0.19
temp_bins: (14.9,16.6]	-0.45	-0.64	-0.26
temp_bins: (16.6,18.4]	-0.55	-0.75	-0.36
temp_bins: (18.4,25.2]	0.10	-0.09	0.29

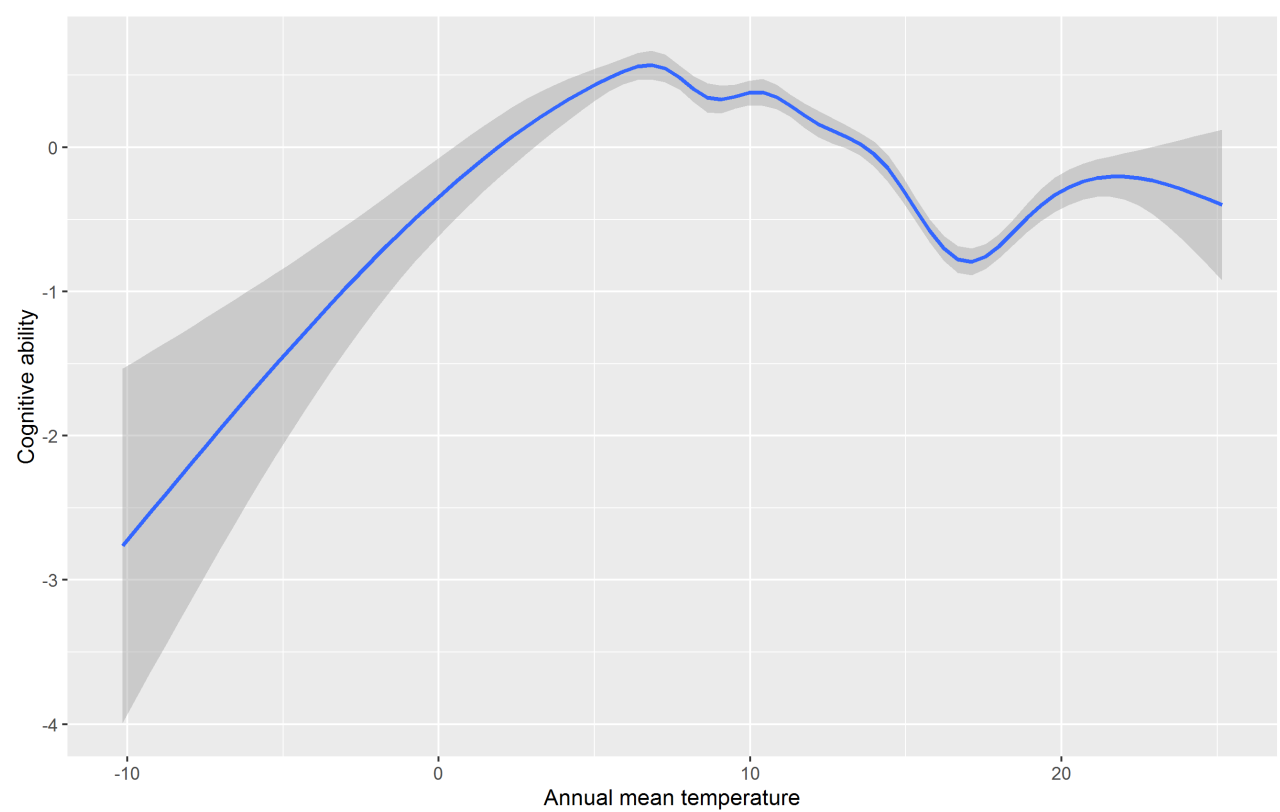


Figure 11: Annual mean temperature and cognitive ability. Line fit by LOESS.

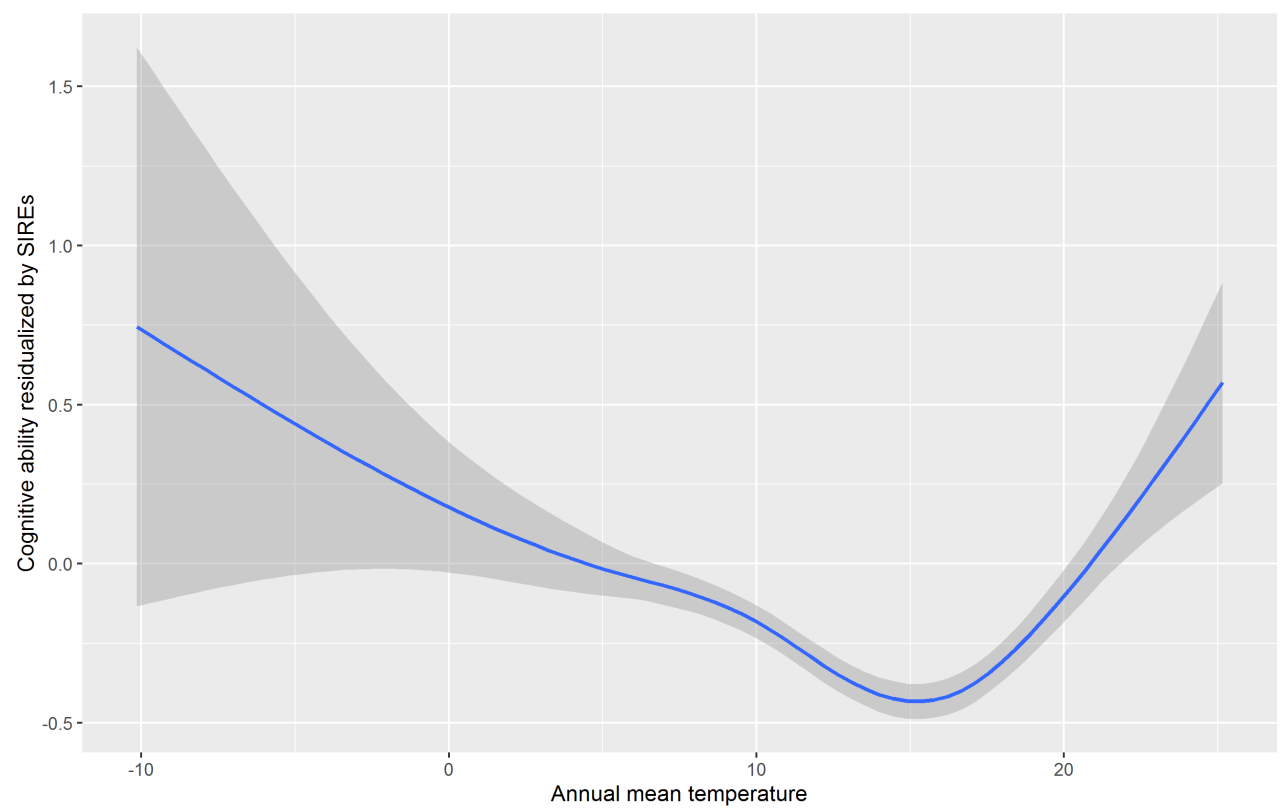


Figure 12: Annual mean temperature and cognitive ability residualized by SIRE model. Line fit by LOESS.

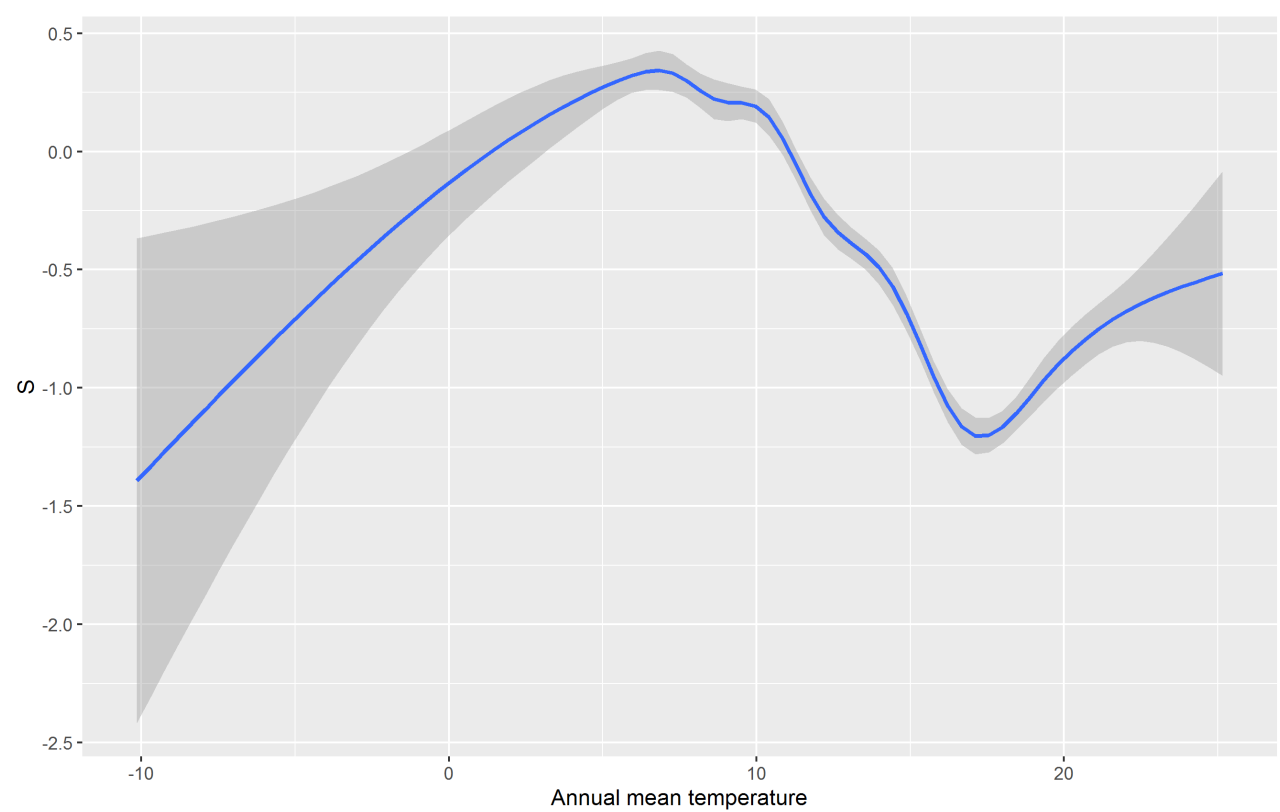


Figure 13: Annual mean temperature and S. Line fit by LOESS.

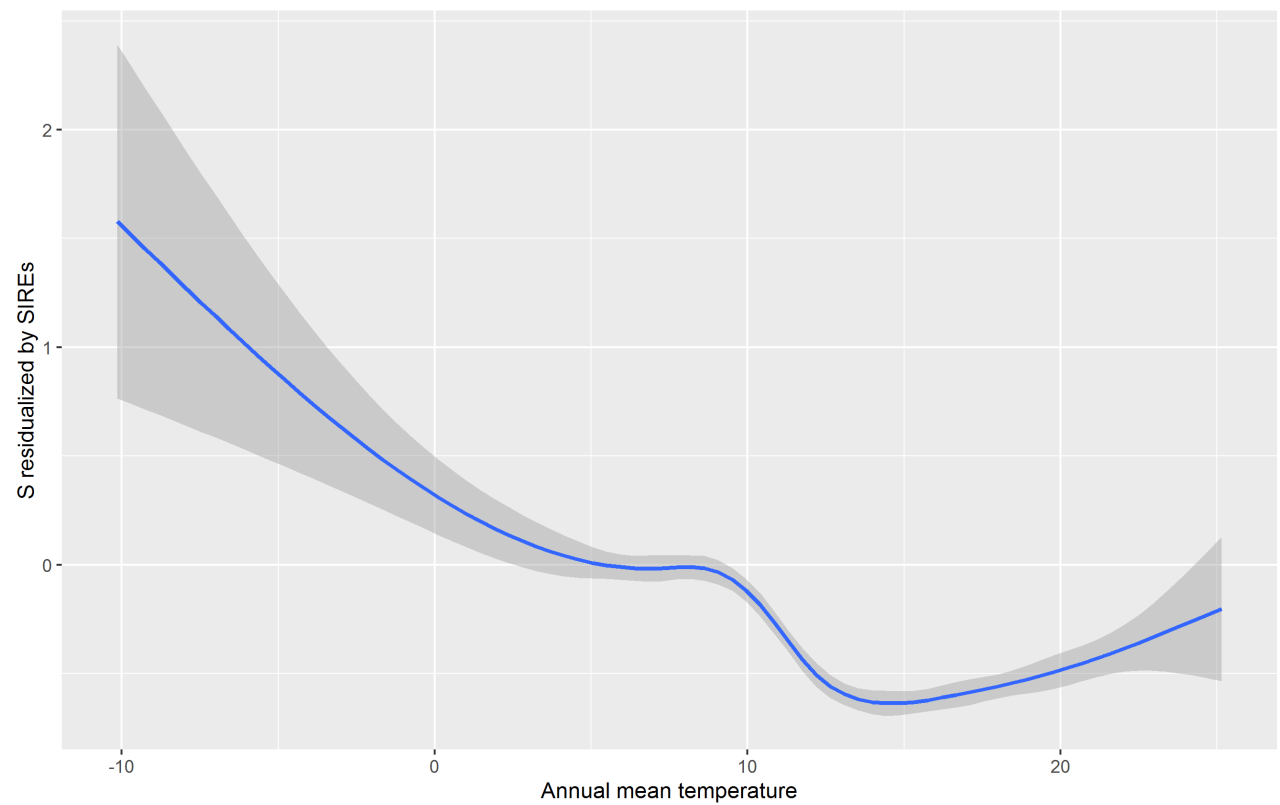


Figure 14: Annual mean temperature and S residualized by SIRE model. Line fit by LOESS.

Table 16: OLS regression betas. Outcome: S. Predictors: SIREs + discretized temperature. Cross- validated R = .71.

Predictor	Beta	CI lower	CI upper
White	0.29	0.26	0.33
Black	-0.28	-0.31	-0.25
Asian	0.21	0.19	0.22
Amerindian	-0.23	-0.28	-0.17
temp_bins: (6.53,8.17]	0.09	-0.10	0.27
temp_bins: (8.17,9.56]	0.04	-0.13	0.20
temp_bins: (9.56,10.7]	-0.07	-0.24	0.09
temp_bins: (10.7,11.9]	0.04	-0.13	0.20
temp_bins: (11.9,13.4]	-0.38	-0.55	-0.21
temp_bins: (13.4,14.9]	-0.36	-0.53	-0.18
temp_bins: (14.9,16.6]	-0.34	-0.51	-0.17
temp_bins: (16.6,18.4]	-0.47	-0.65	-0.30
temp_bins: (18.4,25.2]	-0.32	-0.49	-0.16

We see that the SIRE variables' betas were unchanged in direction, but smaller in size for cognitive ability except for White. The strange non-linear pattern seen in Figure 14 was repeated in the beta values, namely that there was a steep decline in the betas beginning around temperature = 10-12 °C.

5.3 Non-linear effects of all environmental predictors

Given the non-linear effects of temperature, we might wonder about the other environmental predictors. Do they also show non-linear effects? To investigate whether this was the case, all the environmental predictors were discretized into 10 bins in the same way temperature was before and entered into regression models. Tables 17 and 18 shows the model betas.

It can be difficult to get an overview of the importance of variables when presented with every level. Some levels may have strong betas but if few cases fall in those levels, they will not be important in the overall model. To see the importance of each predictor across all levels, one can conduct an analysis of variance and then calculate the eta squared for each. This is the amount of variance (R^2) attributable to that predictor across all levels. Tables 19 and 20 shows this for the model two models.

Table 19: Analysis of variance for regression with discretized environmental predictors. Outcome: cognitive ability. Predictors: SIREs and natural environmental variables.

Predictor	Eta2	Partial Eta2
White	0.36	0.52
Black	0.02	0.05
Asian	0.02	0.07
Amerindian	0.01	0.02
temp_bins	0.07	0.18
lat_bins	0.04	0.11
lon_bins	0.12	0.27
precip_bins	0.00	0.01
elevation_bins	0.01	0.02

Table 17: OLS regression betas with discretized environmental predictors. Outcome: cognitive ability. Predictors: SIREs and environmental predictors. Cross-validated R = .76.

Predictor	Beta	CI lower	CI upper
White	0.39	0.35	0.43
Black	-0.32	-0.36	-0.27
Asian	0.14	0.12	0.15
Amerindian	-0.15	-0.20	-0.09
temp_bins: (6.53,8.17]	0.10	-0.08	0.29
temp_bins: (8.17,9.56]	0.12	-0.06	0.29
temp_bins: (9.56,10.7]	0.24	0.06	0.43
temp_bins: (10.7,11.9]	0.44	0.23	0.64
temp_bins: (11.9,13.4]	0.00	-0.23	0.22
temp_bins: (13.4,14.9]	0.14	-0.12	0.41
temp_bins: (14.9,16.6]	0.15	-0.13	0.44
temp_bins: (16.6,18.4]	0.08	-0.23	0.40
temp_bins: (18.4,25.2]	0.06	-0.29	0.40
lat_bins: (31.7,33.8]	-0.04	-0.19	0.11
lat_bins: (33.8,35.6]	-0.42	-0.60	-0.24
lat_bins: (35.6,37.1]	-0.48	-0.68	-0.29
lat_bins: (37.1,38.4]	-0.57	-0.80	-0.34
lat_bins: (38.4,39.7]	-0.14	-0.38	0.10
lat_bins: (39.7,41]	-0.28	-0.53	-0.03
lat_bins: (41,42.8]	-0.11	-0.38	0.16
lat_bins: (42.8,45]	-0.14	-0.43	0.16
lat_bins: (45,69.4]	0.32	0.01	0.63
lon_bins: (-109,-100]	0.71	0.55	0.87
lon_bins: (-100,-96.7]	1.25	1.08	1.41
lon_bins: (-96.7,-93.6]	1.17	1.01	1.33
lon_bins: (-93.6,-90.4]	0.77	0.60	0.93
lon_bins: (-90.4,-87.2]	0.76	0.60	0.92
lon_bins: (-87.2,-84.5]	0.81	0.64	0.97
lon_bins: (-84.5,-82.1]	0.96	0.80	1.13
lon_bins: (-82.1,-78.2]	1.17	1.02	1.33
lon_bins: (-78.2,178]	1.30	1.14	1.46
precip_bins: (430,603]	0.22	0.11	0.34
precip_bins: (603,800]	0.04	-0.13	0.20
precip_bins: (800,923]	0.19	0.02	0.36
precip_bins: (923,1.03e+03]	0.18	0.01	0.35
precip_bins: (1.03e+03,1.12e+03]	0.23	0.06	0.39
precip_bins: (1.12e+03,1.19e+03]	0.41	0.24	0.58
precip_bins: (1.19e+03,1.28e+03]	0.26	0.10	0.42
precip_bins: (1.28e+03,1.39e+03]	0.18	0.02	0.33
precip_bins: (1.39e+03,4.3e+03]	0.22	0.06	0.37
elevation_bins: (38.9,105]	0.08	-0.02	0.17
elevation_bins: (105,167]	-0.05	-0.15	0.06
elevation_bins: (167,216]	-0.11	-0.22	0.00
elevation_bins: (216,259]	0.00	-0.12	0.12
elevation_bins: (259,317]	-0.03	-0.15	0.09
elevation_bins: (317,396]	-0.07	-0.20	0.06
elevation_bins: (396,572]	-0.19	-0.33	-0.06
elevation_bins: (572,1.01e+03]	0.17	0.01	0.33
elevation_bins: (1.01e+03,3.06e+03]	0.43	0.23	0.63

Table 18: OLS regression betas with discretized environmental predictors. Outcome: S. Predictors: SIREs and environmental predictors. Cross-validated R = .71.

Predictor	Beta	CI lower	CI upper
White	0.32	0.28	0.37
Black	-0.32	-0.36	-0.27
Asian	0.21	0.20	0.23
Amerindian	-0.20	-0.25	-0.14
temp_bins: (6.53,8.17]	0.07	-0.11	0.26
temp_bins: (8.17,9.56]	0.07	-0.11	0.24
temp_bins: (9.56,10.7]	0.08	-0.11	0.27
temp_bins: (10.7,11.9]	0.12	-0.09	0.32
temp_bins: (11.9,13.4]	-0.33	-0.56	-0.10
temp_bins: (13.4,14.9]	0.02	-0.24	0.28
temp_bins: (14.9,16.6]	0.01	-0.28	0.29
temp_bins: (16.6,18.4]	-0.22	-0.54	0.10
temp_bins: (18.4,25.2]	-0.04	-0.39	0.30
lat_bins: (31.7,33.8]	0.28	0.14	0.43
lat_bins: (33.8,35.6]	-0.03	-0.21	0.14
lat_bins: (35.6,37.1]	-0.25	-0.45	-0.06
lat_bins: (37.1,38.4]	-0.33	-0.56	-0.11
lat_bins: (38.4,39.7]	0.25	0.01	0.49
lat_bins: (39.7,41]	-0.10	-0.34	0.15
lat_bins: (41,42.8]	0.00	-0.28	0.27
lat_bins: (42.8,45]	-0.09	-0.38	0.21
lat_bins: (45,69.4]	0.12	-0.18	0.43
lon_bins: (-109,-100]	-0.02	-0.18	0.14
lon_bins: (-100,-96.7]	-0.05	-0.21	0.11
lon_bins: (-96.7,-93.6]	0.13	-0.03	0.29
lon_bins: (-93.6,-90.4]	0.01	-0.16	0.17
lon_bins: (-90.4,-87.2]	0.00	-0.17	0.16
lon_bins: (-87.2,-84.5]	-0.09	-0.26	0.08
lon_bins: (-84.5,-82.1]	-0.09	-0.25	0.08
lon_bins: (-82.1,-78.2]	0.17	0.01	0.33
lon_bins: (-78.2,178]	0.53	0.37	0.69
precip_bins: (430,603]	0.00	-0.12	0.11
precip_bins: (603,800]	0.10	-0.06	0.27
precip_bins: (800,923]	0.20	0.04	0.37
precip_bins: (923,1.03e+03]	0.00	-0.16	0.17
precip_bins: (1.03e+03,1.12e+03]	-0.08	-0.25	0.08
precip_bins: (1.12e+03,1.19e+03]	0.00	-0.16	0.16
precip_bins: (1.19e+03,1.28e+03]	0.02	-0.14	0.18
precip_bins: (1.28e+03,1.39e+03]	-0.10	-0.25	0.06
precip_bins: (1.39e+03,4.3e+03]	0.01	-0.15	0.17
elevation_bins: (38.9,105]	0.02	-0.07	0.12
elevation_bins: (105,167]	0.02	-0.08	0.13
elevation_bins: (167,216]	0.02	-0.09	0.13
elevation_bins: (216,259]	0.04	-0.08	0.16
elevation_bins: (259,317]	0.17	0.05	0.29
elevation_bins: (317,396]	-0.06	-0.19	0.07
elevation_bins: (396,572]	-0.07	-0.20	0.06
elevation_bins: (572,1.01e+03]	-0.11	-0.27	0.05
elevation_bins: (1.01e+03,3.06e+03]	0.32	0.11	0.52

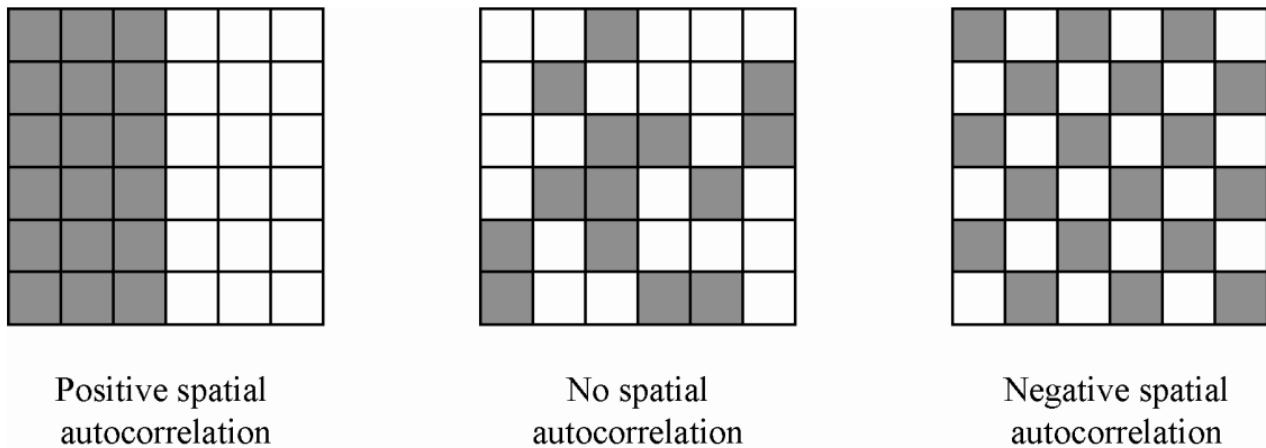


Figure 15: Illustrations of spatial autocorrelation ([Radil, 2011](#))

Table 20: Analysis of variance for regression with discretized environmental predictors. Outcome: S. Predictors: SIRE and natural environmental variables.

Predictor	Eta2	Partial Eta2
White	0.14	0.26
Black	0.17	0.30
Asian	0.20	0.33
Amerindian	0.01	0.02
temp_bins	0.03	0.08
lat_bins	0.02	0.06
lon_bins	0.03	0.06
precip_bins	0.01	0.01
elevation_bins	0.01	0.02

We can now more clearly see that results indicate that White, Black and Asian (the latter two only for S) are important predictors while the environmental variables and Amerindian % were not. The SIRE variables explained most of the variation explained by the models: 62 % / 84 % for cognitive ability and S, respectively. This is the opposite pattern of what was found for the US state-level analysis, where S was more explainable by non-SIRE predictors than cognitive ability ([Fuerst & Kirkegaard, 2016a](#), Section 11).

Because the models now have 49 predictors ($4 + 5 \times 9$), overfitting is a problem. To explore how large a problem, LASSO regression was run with the discretized environmental predictors as well as the primary SIRE variables. No consistent pattern emerged in the results aside from two: First, the SIRE predictors continued to be useful and have the same signs, and second, occasionally, one bin of some environmental variable was useful, while others were not, even its own neighbors. I cannot think of a sensible causal model that can explain why these bins would be important but the neighboring bins would not. It seems more likely to be some spatial confound.

The full results can be found in the supplementary materials (sheets `envir_nonlin_CA_lasso` and `envir_nonlin_S_lasso`).

6 Spatial patterns in the data

Administrative divisions are units that have a spatial position. Usually, this means that cases that are located close to each other are similar, a phenomenon known as positive spatial autocorrelation. Figure 15 illustrates the concept.

Standard statistical methods assume that error terms are not correlated in any meaningful way, but with spatial autocorrelation in the data, the error terms (residuals) of nearby cases can become correlated, which has the effect of biasing some model estimates ([Gelade, 2008](#); [Hassall & Sherratt, 2011](#)).

[Hassall & Sherratt \(2011\)](#) proposed the following strategy for handling the problem:

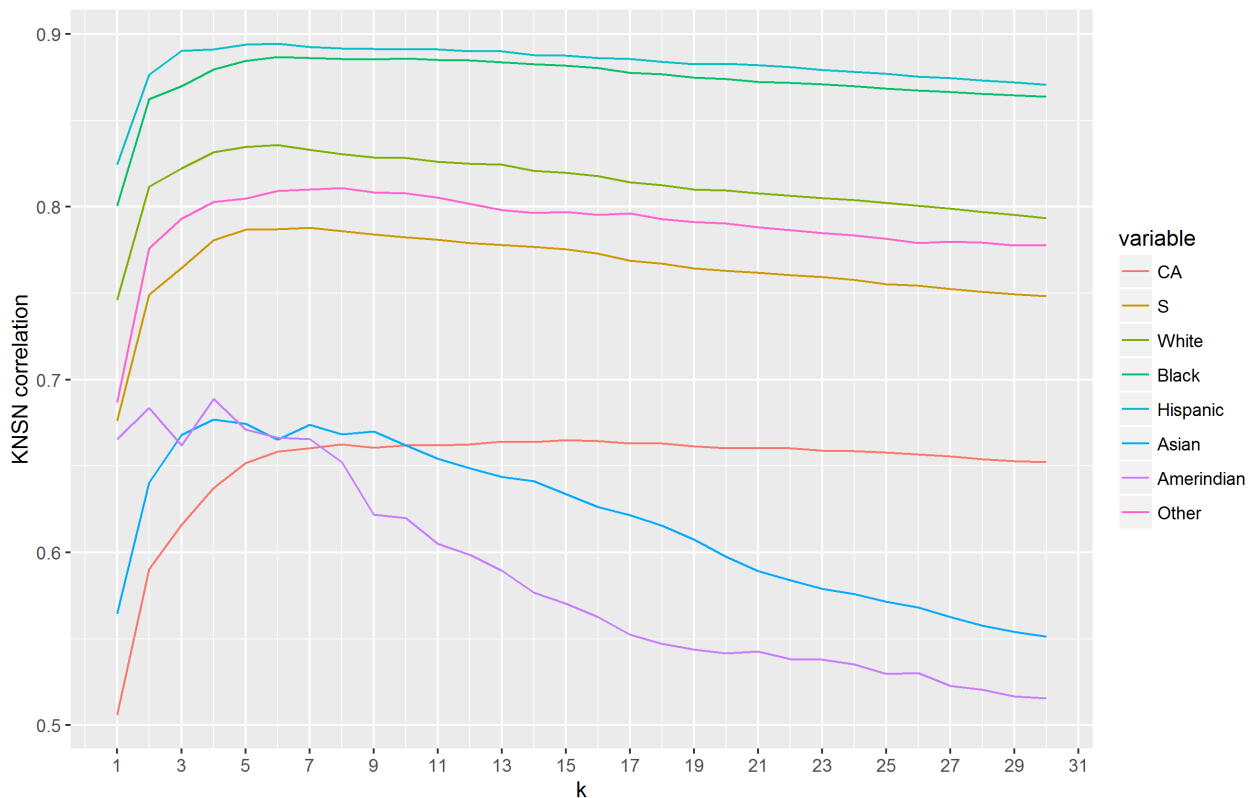


Figure 16: k-nearest spatial neighbor regression correlations for cognitive ability, S and SIREs by k.

1. Calculate spatial autocorrelation measures for the primary variables.
2. If step 1 indicates a sizable amount of spatial autocorrelation in the primary variables, then fit the models of interest and save the residuals. Calculate spatial autocorrelation measures for the residuals.
3. If step 2 indicates that there is a sizable amount of spatial autocorrelation in the residuals, then control for spatial autocorrelation using specific spatial models.

Fuerst & Kirkegaard (2016a) used this approach and it was also used in this study.

6.1 Measured spatial autocorrelation in the data

The latitude and longitude of each county was available and with those one can calculate measures of spatial autocorrelation (SAC). The most commonly used method is Moran's I, but there are also others such as Geary's C, the Mantel test and k-nearest spatial neighbor regression (KNSNR) correlation (Radil, 2011; Kirkegaard, 2015d). Of these, only the KNSNR correlation has a simple intuitive meaning: it is the correlation of the predicted values for cases based on their k-nearest neighbors with the observed values for the same cases. It is a direct measure of how well one can predict the value of a given case by only looking at its neighbors.

The standard implementation of Moran's I is a universal measure of spatial autocorrelation. KNSNR, on the other hand, is always local. One must choose how local, by choosing a value of k; the number of neighbors to use in the predictions. The value of k specifies the reach of the neighbor effects. To estimate what a good value of k would be, the KNSNR correlations were calculated for every k = 1-30 for cognitive ability, S and the SIRE variables. Figure 16 shows the results.

As was observed in the other studies, a value of k around 3-10 produces the largest correlations. In this case, k=7 had the strongest results (mean r = .78), but the differences were trivial (k = 4-10 all had mean r's ≥ .77). Nevertheless, the following analyses uses k=7.

I have implemented three of the SAC measures in a convenient function for R and this was used to measure SAC for each variable. Table 21 shows the results.

Table 21: Measures of spatial autocorrelation in the primary variables.

Variable	Moran's I	Mantel test	KNSNR 7
CA	0.10	0.08	0.62
S	0.16	0.07	0.76
White	0.14	0.17	0.82
Black	0.20	0.02	0.87
Hispanic	0.15	0.20	0.89
Asian	0.06	0.24	0.67
Amerindian	0.03	0.22	0.66
Other	0.08	0.25	0.79

To my knowledge, there are no useful guidelines for the effect sizes of Moran's I or the Mantel test. However, KNSNR is a correlation, so the usual guidelines apply. We see that using the counties' neighboring cases, the correlations range from .62 to .89, all of which are large by conventional standards. In other words, by knowing the values of the neighbors of a given county, one can predict the values of that county fairly well.

6.2 SAC in model residuals

Since the primary variables are strongly SAC, SAC should be measured in the model residuals as well. If these are also strongly SAC, then there are spatial patterns in the outcome variables (cognitive ability and S) that cannot be accounted for by those in the predictor variables. To check this, I fitted the best SIRE model from before for both cognitive ability and S and saved the residuals. After this, the residuals were examined for SAC. Table 22 shows the results with the results for the original outcome variables for comparison.

Table 22: Spatial autocorrelation in outcome variables before and after residualization by the SIRE model.

Variable	Moran's I	Mantel test	KNSNR 7
CA resids	0.07	0.02	0.60
S resids	0.13	0.05	0.79
CA	0.10	0.08	0.66
S	0.16	0.07	0.79

Unexpectedly, we see that the residuals show about the same level of SAC as do the original values. Interpreting this is tricky. It does not mean that the predictors could not explain the SAC in the outcomes variable at all. Because if predictors could only predict the non-SAC variance in the outcome, then SAC would increase in the residuals. Thus, when the amount of SAC is unchanged, it seems to me that this means that the predictors were equally good at predicting the SAC and non-SAC variance in the outcome. This inference should be tested in a future simulation study.

6.3 Controlling for SAC

Since there was strong SAC in the model residuals, the model assumptions underlying OLS regression are violated (correlated errors), and the predictors may be spurious causes that happen to be SAC with the true cause(s). We can try to correct for this by using an explicit spatial model. There are multiple ways to explicitly control for SAC which rely on different assumptions (Diniz-Filho et al., 2009). Hassall & Sherratt (2011) use a method called spatial eigenvector mapping. Unfortunately, they used an external application to perform this analysis and I have not been able to find a suitable implementation in R. To control for SAC, I developed two other methods: one based on KNSNR and one based on spatial local regression (SLR) (Kirkegaard, 2015d).

The idea with the KNSNR correction is to first predict the values of each case using the neighbors' values. Then, one can either enter these as a predictor in a model along with the predictors one's interested in ('multiple regression variant'), or use them to partial out the spatial pattern in the dependent variable and use the residuals as the new dependent variable ('partial correlation variant'). The first variant can be problematic when SAC is

strong and the predictors are good at predicting the dependent variable because this means that the model will have a high multicollinearity.

The idea with SLR is combine the ideas of spatial statistics and local regression. Since the SAC has to do with the overall patterns in the data, one can effectively minimize the problem by looking at very small subsets of the data. Specifically, for each case, one finds its $k-1$ nearest neighbors. One then uses the k cases (the case and its $k-1$ neighbors) to fit the model and saves the results. This is done for every case. The results are then aggregated (akin to meta-analysis) in the end to produce an overall SAC- minimized estimate.

For the methods, $k=7$ was chosen in accordance with the findings in Section 6.1. The partial correlation method was employed, i.e. all variables had SAC partialled out. This method is more conservative in that it probably results in some true predictor variance being removed from the outcome variable. Tables 23 and 24 show the betas for the SIRE model with and without controls for SAC for each outcome variable.

Table 23: Model betas pre- and post-SAC correction. Outcome: cognitive ability.

Predictor	Uncorrected	KNSNR	SLR	Mean controlled
White	0.36	0.46	0.35	0.41
Black	-0.32	-0.11	-0.16	-0.13
Asian	0.15	0.18	0.29	0.24
Amerindian	-0.16	-0.05	-0.08	-0.06

Table 24: Model betas pre- and post-SAC correction. Outcome: S.

Predictor	Uncorrected	KNSNR	SLR	Mean controlled
White	0.35	0.49	0.35	0.42
Black	-0.37	-0.19	-0.16	-0.18
Asian	0.33	0.22	0.41	0.32
Amerindian	-0.12	-0.12	-0.13	-0.12

Generally, the betas were fairly unchanged by the SAC control procedures, with the exception of Black which, while it kept its negative sign, was substantially reduced for both outcome variables and across both methods.

Assuming that the methods work, this indicates that the relationship between the predictors and the outcomes were not confounded by unobserved SAC causes.

7 Multilevel analyses

Counties are located within states which, in turn, are located inside the US. This hierarchical nature of the data can result in correlated errors for the same reason that SAC can.⁶ This can be accommodated by using models designed to analyze such data variously called *multilevel* or *hierarchical* models. Unfortunately, multilevel models require a fairly large sample size to fit and I was unable to fit true multilevel models to the present data. However, it is still possible to carry out multilevel analyses in the sense that one can analyze features of the dataset within each state and at the state-level by aggregating the data inside each state (*'pseudo-multilevel'*).

7.1 Stability of the S factor structure across levels

It is possible that when subsets of the data are examined, there is no general factor at all; that the S factor that appears when all the data are analyzed is due to state-level effects. There are three ways one can examine the data:

⁶ Of course, since the counties of a given state are all located in a particular area of the country, state-related correlated errors may simply be due to SAC.

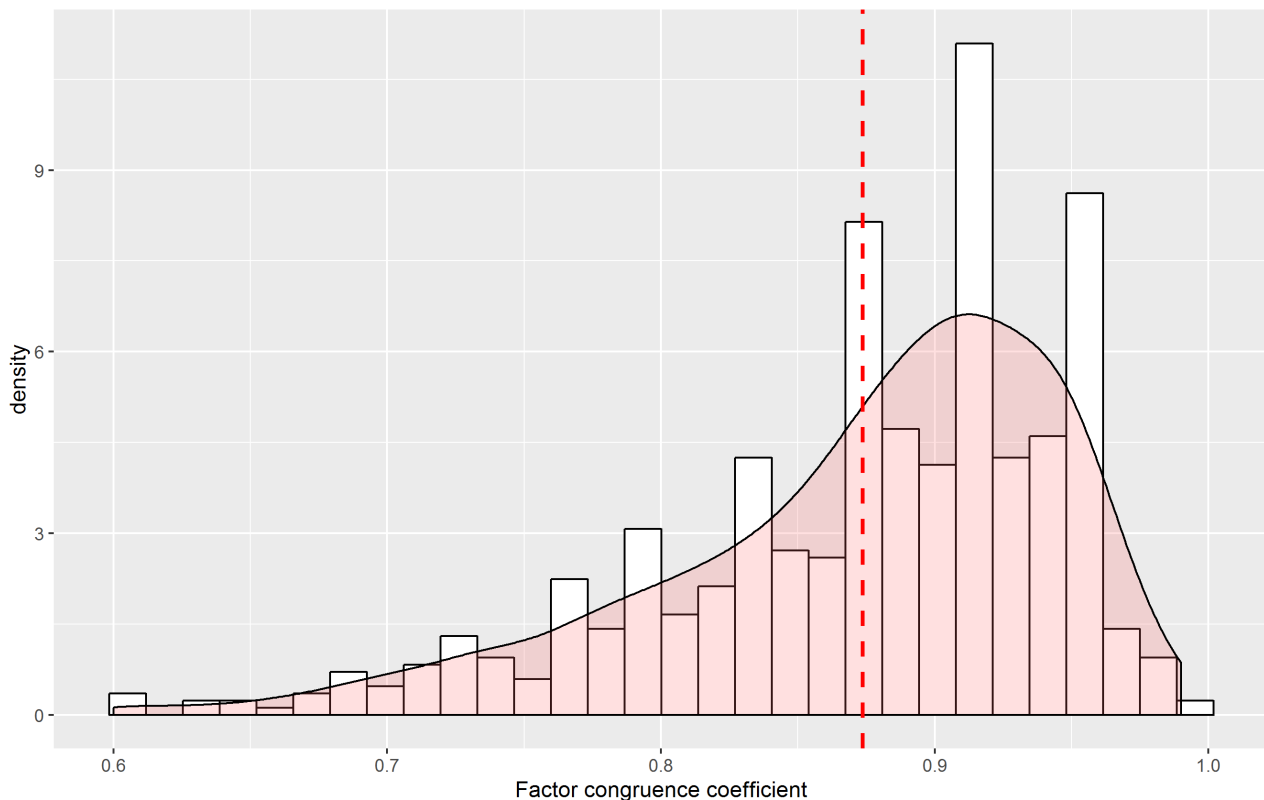


Figure 17: Density-histogram of congruence coefficients. The vertical line is the mean value (.87). N=630

1. Analyze all the counties in one analysis. This is what was done in Section 3.4.
2. Split the counties into sets based on states and run the factor analysis in each set.
3. Split the counties into sets based on states, calculate an average of each S indicator within each state and then factor analyze the data. This would be the same as analyzing the data at the state- level.

7.1.1 Factor analysis of counties by state

Because some states had too few cases for factor analysis to work, only the states where there were no extraction errors were included (N=37). The loadings of each factor analysis were then saved and compared using congruence coefficients (Lorenzo-Seva & ten Berge, 2006). Figure 17 shows a density-histogram of the congruence coefficients.

As can be seen, the factor congruence coefficients were very strong, especially considering that many states only have a few counties (e.g. Oregon has 35). Thus, if one were to somehow control for the 'sampling error'⁷, the factor structure would possibly be perfectly identical across states. I don't know any method to do this, but one could try to confirm the diagnosis by analyzing the deviation from the main analysis and the number of cases for each analysis. To do this, the mean absolute difference in factor loadings from the main analysis were calculated, and then correlated it with the number of counties by state. Figure 18 shows the results.

There is indeed a negative correlation such that the states with more counties deviate less from the results using all the data. An alternative method was also tried using congruence coefficients compared to the main model. This gave nearly identical results ($r = .39$). The two measures correlated at $-.92$. Thus, it appears that the factor structure was very similar across levels and the dissimilarity that was observed was due at least in part to the small number of cases in some analyses.

⁷ I note that talking of *sampling error* assumes that we are trying to infer features of some larger population. Of course, in one sense we already have the entire population of counties; there are no more counties out there to add to the dataset. However, in another sense, there could have been more and if there were more, natural random variation would not affect the similarity of factor structure so much.

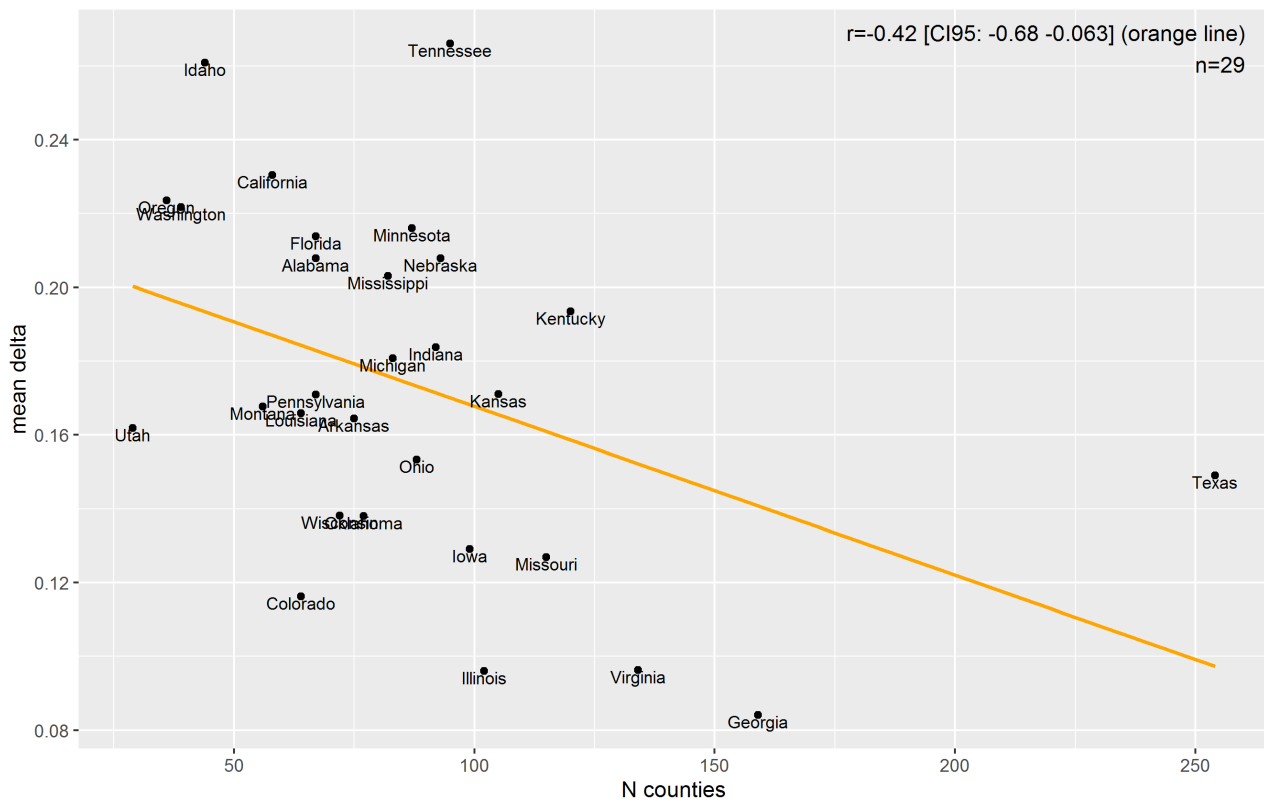


Figure 18: Difference to the full analysis loadings and number of counties by state.

7.1.2 Factor analysis at the state-level

There has already been published two large S factor analyses of US states ([Kirkegaard, 2015c,a](#)), so carrying out another may seem redundant. However, they used different variables, and the structure at the state-level may not be similar to that found when all the counties are examined together. To determine the similarity of the S factor structures at the two levels, I calculated the mean of each S indicator within each state, and then factor analyzed the aggregated data. Figure 19 shows the results with the factor analysis from the main county-level analysis for comparison.

The loadings are nearly identical across levels. The congruence coefficient is .98. A similar result was seen in a previous unpublished study of a general religious factor among Muslims ([Kirkegaard, 2015b](#)).

7.2 Multilevel regression with SIREs

In Section 4.1, we saw that in the best model, White and Asian had positive betas while Black and Amerindian had negative betas. It is possible, though, that these values arise from some aggregation effect. To examine whether this was the case, the SIRE model was fit within each state. Table 25 shows the summary statistics for the model betas.

Table 25: Summary statistics for SIRE predictors for models fit by state.

Statistic	White	Black	Asian	Amerindian
mean	-0.30	-1.46	0.61	-3.25
median	0.37	-0.63	0.58	-0.68
sd	2.55	4.59	0.87	7.72
mad	0.61	1.01	0.45	1.56
fraction zero / NA	0.02	0.02	0.04	0.06

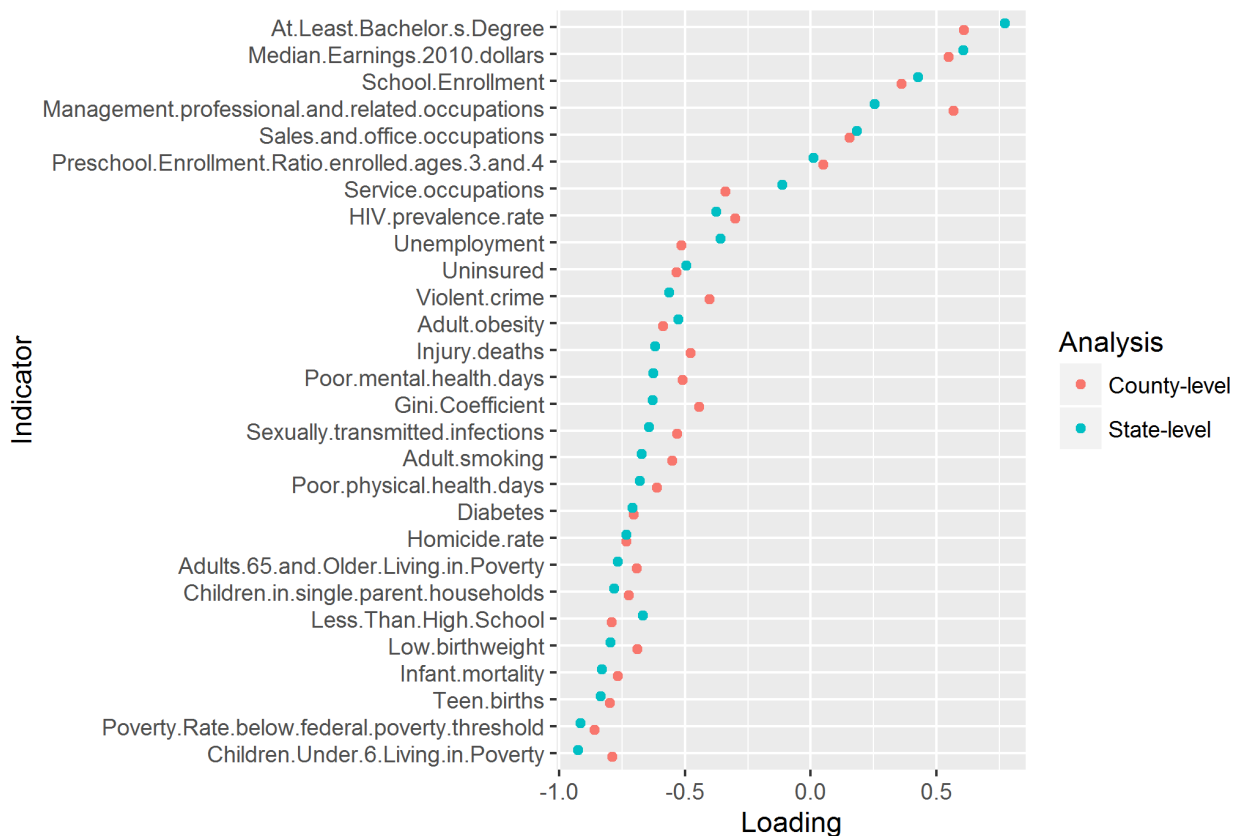


Figure 19: Factor loadings at the county- and state-level.

Due to the small (too small) size of some states in terms of counties, some nonsensical betas were obtained.⁸ These throw off the standard measures (mean and sd), so it is important to focus on the robust statistics (median and mad). These show that the usual pattern of results were found when the data were analyzed in this way. If anything, they were somewhat stronger, indicating that state-level effects were suppressing the betas somewhat.

8 Fixed effects of states

During the review, Noah Carl suggested including dummies for the states into the multiple regressions. To do this, three regressions were run, all of which had S as the outcome variable. In the first, cognitive ability and State were included as predictors. In the second, the SIRE predictors were added, and in the third, the discretized natural environmental predictors were added as well. Due to the large number of predictors and their levels (89 in the third model), it would take too much space to present the full table of predictor betas. Instead, the ANOVA-based eta squared values are given in Tables 26, 27 and 28. The full results can be found in the supplementary materials (sheets State_fixed_1, State_fixed_2 and State_fixed_3).

Table 26: ANOVA-based effect sizes. Outcome: S. Predictors: cognitive ability, state. Cross-validated R = .83.

Predictor	Eta ²	Partial Eta ²
CA	0.47	0.64
State	0.27	0.50

⁸ For instance, one state had a negative standardized beta of -12 for White %, while another had a positive beta of 21 for Amerindian %.

Table 27: ANOVA-based effect sizes. Outcome: S. Predictors: cognitive ability, state, SIREs. Cross-validated R = .86.

Predictor	Eta ²	Partial Eta ²
CA	0.47	0.68
White	0.00	0.02
Black	0.09	0.30
Asian	0.13	0.37
Amerindian	0.00	0.01
State	0.08	0.27

Table 28: ANOVA-based effect sizes. Outcome: S. Predictors: cognitive ability, state, SIREs, discretized natural geospatial and climatological variables. Cross-validated R = .86.

Predictor	Eta ²	Partial Eta ²
CA	0.47	0.71
White	0.00	0.01
Black	0.10	0.34
Asian	0.14	0.42
Amerindian	0.00	0.01
State	0.08	0.30
lat_bins	0.01	0.04
lon_bins	0.00	0.02
precip_bins	0.01	0.04
temp_bins	0.01	0.03

As can be seen, the three models had approximately equal predictive power (multiple R's .83 to .86) and this was mostly due to the predictive power of cognitive ability (47 % of the variance in all models, corresponding to R = .69). Without the SIRE predictors, State had substantial incremental validity (27 % of variance; R = .52), but when SIREs were included, it was reduced to a medium to small effect (8 % variance; R = .28). The environmental predictors had only trivial effects and did not detectably increase the R² when using cross-validation.

9 Does cognitive ability mediate the relationship between SIREs and S?

The central tenet of cognitive sociology is that cognitive ability is an important causal variable to explain variation in socioeconomic outcomes for individuals as well as groups of individuals. A stronger version of this claim is that cognitive ability explains all the differences in socioeconomic outcomes. This view is not tenable at the individual-level because there are known non-cognitive-ability traits that differ between individuals and (seem to) cause differences in socioeconomic outcomes. These traits include conscientiousness which includes facets of orderliness, self-discipline/self-control and achievement striving; extroversion which includes the facets of assertiveness and gregariousness; neuroticism/emotional stability which is strongly related to mental health; and interests which have an important influence on which education and jobs people seek (Bogg & Roberts, 2004, 2013; De Fruyt & Mervielde, 1999; Dudley et al., 2006; Lubinski, 2000; Roberts et al., 2007; Woods & Hampson, 2010).

Just because a trait explains differences in outcomes between individuals does not mean that it also explains differences in outcomes between groups. For this to happen, the groups must also differ in the trait in question. Men and women, for instance, differ substantially in some aspects of personality and in occupational interests and which is likely a large part of the reason why they differ in job types (Costa Jr. et al., 2001; Giudice et al., 2012; Lippa, 2010; Su et al., 2009; Woods & Hampson, 2010). Del Giudice et al. (2012) reported a .89 d for warmth (favoring women) and a .54 d for dominance (favoring men); Su et al. (2009) reported a .93 d for preference for working with people over things (favoring women).

With regards to SIRE groups, it is well-known that they differ in average levels of cognitive ability (Fuerst, 2014; Roth et al., 2001), but it is less known whether and if so, how much they differ in other traits. A very large meta-analysis found that there are only trivial differences in the OCEAN (big five) personality traits between Whites and Blacks in the US (all d 's < .10; (Tate & McDaniel, 2008)). On the other hand, Schmitt et al. (2011) found White-Black differences of .23 to .33 on five of the six RIASEC scales (see also Walker & Tracey (2012)). Based on this limited review, it seems that cognitive ability may be more important than the other traits in explaining differences in socioeconomic outcome for Whites and Blacks. There is currently insufficient evidence to say whether this is also the case for other observed differences, e.g. Whites vs. Asians.

In our previous study (Fuerst & Kirkegaard, 2016a, Section 13); but see also (Pesta, 2016; Fuerst & Kirkegaard, 2016b) we examined whether (rough estimates of) genomic racial (bio-geographical) ancestry had predictive validity beyond (rough estimates of) cognitive ability. In general, we found only weak and inconsistent evidence of incremental validity of the genomic ancestry when cognitive ability was taken into account. These analyses, however, relied on national- and state-level data and it may be that the effects are not (easily) detectable at that level of aggregation. For this reason, it is worth examining this question at the county-level as well.

Three methods were employed to examine whether cognitive ability mediated the relationship between SIREs and S:

1. A simple method where cognitive ability is regressed out of the SIRE variables and the residualized SIRE variables are then correlated with S (semi-partial correlations). This was the primary method employed in the previous studies.
2. OLS regression with cognitive ability and the SIRE predictors after cognitive ability has been regressed out of them. The rationale for this method is that the effects of the SIREs seem more clear when analyzed simultaneously rather than individually. This method was not used previously because we did not find evidence of incremental validity of using more than one demographic predictor. In a comparative model, the (unresidualized) SIREs are entered along with cognitive ability with S as the outcome.
3. Path analysis where a number of models of the causal paths are tested.

9.1 Semi-partial correlations

Table 29 shows the semi-partial correlations along with the raw correlations.

Table 29: Semi-partial weighted correlations between residualized SIREs and S. Numbers above the diagonal are the original variables, below the diagonal the residualized variables.

	S	CA	White	Black	Asian	Amerindian
S		0.68	0.34	-0.52	0.33	-0.14
CA	0.71		0.59	-0.40	-0.10	-0.13
White	0.16	0.00		-0.47	-0.49	-0.05
Black	-0.23	0.00	-0.46		-0.05	-0.09
Asian	0.20	0.00	-0.34	0.06		-0.08
Amerindian	-0.02	0.00	-0.20	-0.27	0.01	

It can be seen that the residualization reduced the correlations by about half except for Amerindian which was reduced to near-zero levels. This indicates that much, but not all, of the predictive validity of the SIRE predictors comes from their relationship to cognitive ability, consistent with a partial mediation model.

9.2 Multiple regression models

Table 30 shows the model betas for the best SIRE model with cognitive ability included as a predictor in two versions:

Table 30: Model betas for two models. 1: Outcome: S. Predictors: SIREs and cognitive ability. 2: Outcome: S. Predictors: residualized SIREs and cognitive ability. Model cross-validated R's = .78.

Predictor	Beta	CI upper	CI lower
<i>Raw SIRE predictors</i>			
CA	0.46	0.43	0.49
White	0.18	0.15	0.22
Black	-0.22	-0.25	-0.19
Asian	0.27	0.24	0.29
Amerindian	-0.04	-0.07	-0.02
<i>Residualized SIRE predictors</i>			
CA	0.71	0.68	0.73
White	0.15	0.12	0.18
Black	-0.19	-0.21	-0.16
Asian	0.27	0.24	0.29
Amerindian	-0.04	-0.07	-0.02

Except for cognitive ability itself (.46 vs. .71), the model betas were very similar. Note that one cannot estimate the total R of a model by summing the predictor betas. This is because predictors may be correlated and if they are, their validity does not just add up. One can get the independent predictive validity of the predictors by converting the linear model to an ANOVA and then calculating the eta squared values. Doing so reveals that cognitive ability explained 4.2 times as much variance as did the SIRE predictors combined.

9.3 Path models

Three path models were fitted:

1. A pure mediated model. The effects of SIREs on S are 100 % mediated by cognitive ability.
2. A mediation model with heterogeneous alternative paths. The SIREs are allowed to have an effect on S aside from their effect thru cognitive ability.
3. A mediation model with homogeneous alternative paths. The SIREs are allowed to have an effect on S aside from their effect thru cognitive ability but this is thru a common latent variable. This model essentially stipulates that whatever it is that explains the relationships of SIREs to S that is not thru cognitive ability, but a single other construct.

Figure 20 shows the fitted paths.

The paths found generally mirror the earlier results. In the first model, the SIREs have their usual signs but White is stronger than normally and Black and Asian weaker. The S~CA path also has a familiar value (.68, the same as the weighted correlation). In the second model, Asian has a notably strong non-CA path to S in line with previous results suggesting that Asians do better than one would expect based on their cognitive ability (Herrnstein & Murray, 1994). White has a small positive non-CA effect and Black a moderate negative non-CA effect. In the third model, S is roughly equally caused by the hypothesized non-CA construct and CA which together explain about half the variance in S. The effect of White is primarily thru CA while that of Asian is almost exclusively thru non-CA. The relationship between Black and S is also mostly thru the hypothetical construct.

Model fit statistics, however, indicated that the last two models had perfect fit and the first model had a bad fit according to RMSEA (.38) but decent according to the good fit index (GFI; .990). This probably indicates that there was a technical problem in fitting the models to the data or that they did not fit the data very well. This makes it difficult to interpret the results.

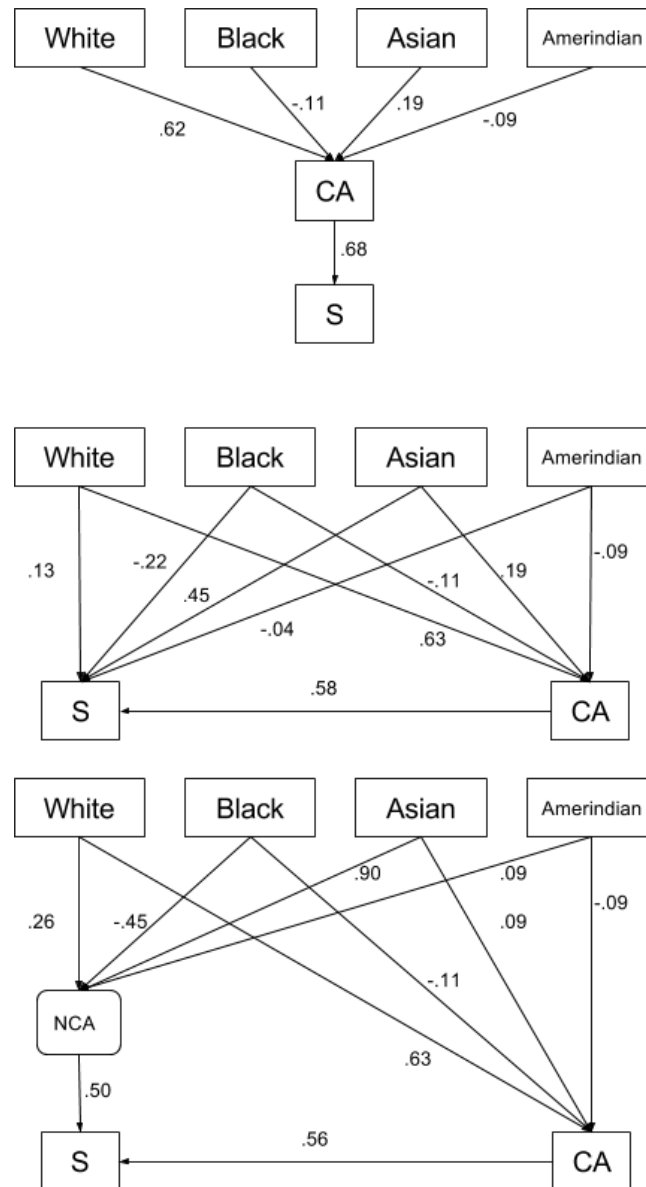


Figure 20: Three path models for modeling the effects of SIREs and cognitive ability on S.

10 Jensen's method

When one analyzes data related to latent variables and one has multiple indicators of the latent variables, it is possible to perform a robustness check using Jensen's method (method of correlated vectors, (Jensen, 1998)). This method consists of correlating the factor loadings of the indicators with the indicators' relationships to the criterion variable. The simple theory behind this procedure is that if the latent variable is really related to the criterion variable, then the indicators that measure the latent variable better should be more strongly related to the criterion variable. To be sure, this method has problems such as that it requires a fair number of indicators which must themselves be representative of the domain. Furthermore, the method is not always able to distinguish between alternative models (Ashton & Lee, 2005). In my opinion, it remains an interesting tool but with unclear interpretation for values that are not close to 1 or -1. Note that due to the existence of indicators with negative loadings, a few modifications have been made to the method which are explained in Kirkegaard (2016b).

In the present study, the S factor was treated as a latent variable measured by 28 indicator variables. The criteria variables are cognitive ability (not treated as a latent variable in this study) and the SIRE variables. Figure 21 shows the relationship between S factor loadings and the indicators' correlations to cognitive ability.

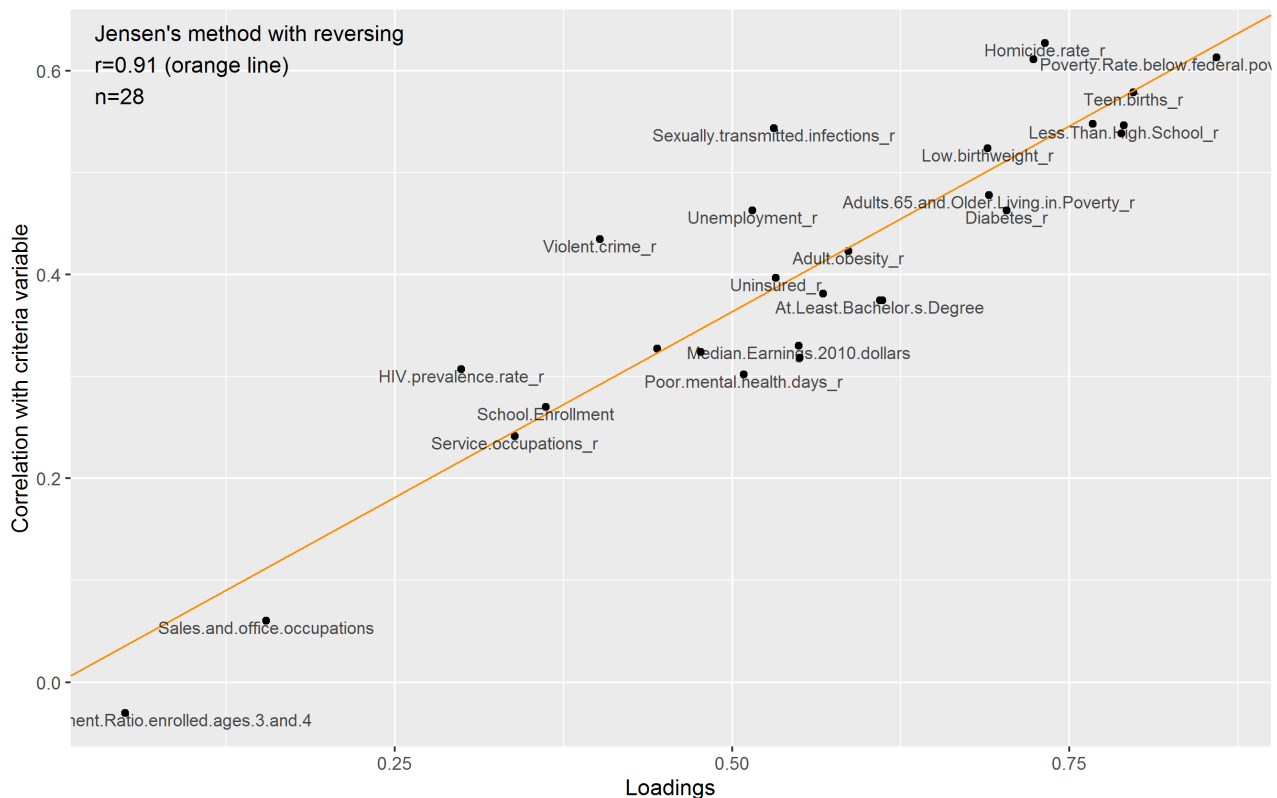


Figure 21: Jensen's method for S and cognitive ability. Note that negative loadings are reversed.

We see that there is a strong relationship between the indicators' relationship to cognitive ability and their loading on S, indicating that cognitive ability is related to the latent S factor itself and not just the other variance.

Using Jensen's method on the SIRE variables is more tricky because one must decide whether just to use the zero-order correlations or the regression beta weights. In view of the finding that the effects of the SIRE predictors was to some degree independent, it seems best to use the regression method. To do this, every indicator was regressed on the four SIRE predictors (White, Black, Asian, Amerindian) and the beta values saved. These values were then correlated with the S loadings of each indicator. Figure 22 shows the results.

We do not see strong correlations altho there are slight correlations in the expected directions (r 's = .10, -.17, .34, -.17). To see whether this near-zero finding was due to the use of a novel method, I also used the simpler method of correlating each indicator with the SIRE predictors. This gave slightly stronger results but still not close to those seen for cognitive ability (r 's .23, -.45, .24, -.12). These findings seem to show that there are substantial relationships between the SIREs and the S indicators that are not just related to the S factor as argued by e.g. (Pesta, 2016).

11 Discussion and conclusion

The results indicated that there is a clear S factor in the US at the county-level. This finding was very robust to variations in methods and controls. As expected, the S factor had a strong correlation to cognitive ability (.68, weighted correlation) as has been found in both the US at the state-level, in many other countries and at the national-level.

With regards to the relationships between demographics and outcomes, the results were not surprising. The more Whites and Asians in a county, the higher the cognitive ability and S of that county. Reversely, the more Blacks, Hispanics or Amerindians, the lower. These are the expected findings based on the already existing knowledge of individual-level, county-level, state-level and country-level outcomes (Beaver & Wright, 2011; Fuerst & Kirkegaard, 2016a,b). Perhaps unexpectedly, SIRE homogeneity was found to have non-linear but generally negative effects in both single (correlation) and multiple regression analysis. This is due to counties

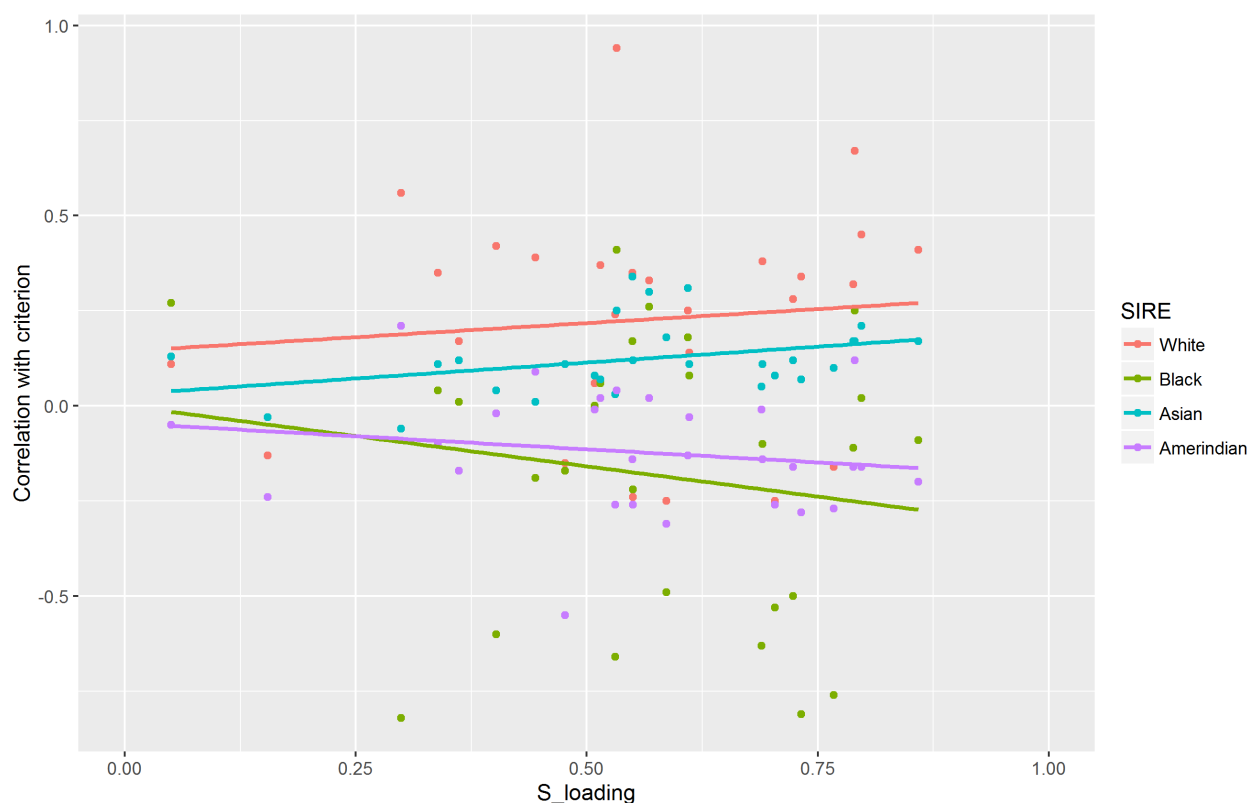


Figure 22: Jensen's method for SIRE predictors and S.

with very high White % and medium to low S scores. The explanation for this pattern is probably related to brain drain and intra-White regional differences (Fischer, 1989).

The present study relied on self-identified race/ethnicity, not genomic estimates of racial (biogeographical) ancestry. If one is interested in the effects of racial ancestry on S, this has several problems. First, some SIRE groups consist of persons of mixed ancestry. For instance, Hispanics are of (mostly Iberian) European, Amerindian and sometimes African ancestry (Bryc et al., 2015). Second, members of the same SIRE group living in different areas vary in their genomic ancestry. For instance, Blacks living in Alabama have less European ancestry than Blacks living in Washington (17 % vs. 30 %, based on 23andme.com customer data; (Bryc et al., 2015, Table S2)). A previous study has shown that within SIRE variation in racial ancestry is related to outcomes in mostly consistent ways at the state-level (Kirkegaard & Fuerst, 2016). It may be possible to obtain county-level genomic estimates in the future that could further test genetic causal models of social inequality.

Multiple natural environmental predictors were examined, but in general they did not add much incremental validity. This is despite the fact that the models allowed for non-linear effects of these. The odd patterns revealed by the non-linear models may reflect the fact that the natural environmental predictors are confounded with other variables.

The spatial control analyses found that the SIREs were not spurious causes. There is some simulation evidence that suggests that these methods work in the sense that when used on relatively simple datasets where SAC has been added, the control methods are able to distinguish between true and spurious causes (Kirkegaard, 2015d). However, these methods are experimental and in need of more extensive testing. For this reason, these results should not be seen as final for this dataset. Others are encouraged to analyze the compiled dataset further.

Multilevel analyses generally did not reveal any surprises. The S factor was very robust to being analyzed within each state and at the state-level. There was some evidence that the differences to the main analyses were due to 'sampling error'. The SIRE model was also found to be robust when analyzed within each state.

Analyses of the fixed-effects of states found that these had little incremental validity once SIREs were included in the models. Other analyses that find effects of states may in fact be finding cryptic effects of SIREs.

Is the relationship between SIREs and S mediated by cognitive ability? This question was examined with a variety of methods, which also resulted in a variety of results spanning from "yes, mostly" to "perhaps half".

Previous analysis revealed that statistically controlling for the effects of SIREs is difficult and that environmental effects that appear in such models can be overestimated (Kirkegaard & Fuerst, 2016). It is possible to get around this problem if one could obtain socioeconomic and cognitive ability at the county level that is broken down by SIRE group. Such data may be available in the census and if so, should be used in a future study.

The application of Jensen's method revealed that the relationship between cognitive ability and the S scores was probably due to the latent S factor. However, it also revealed that the relationship between the S scores and the SIREs were not strongly attributable to the latent S factor. In other words, after equalizing for S, there remains substantial SIRE differences in S indicators. This was also found by Pesta (2016).

The school-district level dataset used in the present study has a much larger sample size (N=17,886). It contains demographic (SIRE) data, but not many socioeconomic variables. If one could obtain them, it may be a good dataset to analyze in a future study.

It should be remembered that statistical patterns may be very different across levels and for this reason, one must be cautious about inferring patterns between levels (ecological fallacy). In general, however, results from analyses at many different levels tend to agree for this area of research.

Supplementary material and acknowledgments

The peer review thread can be found at <http://openpsych.net/forum/showthread.php?tid=274>. The data, R analyses code and high quality figures can be found at <https://osf.io/cknjr/>.

Thanks to Random Critical Analysis (pseudonym, blog <https://randomcriticalanalysis.wordpress.com/>) who helped locate datasets.

Thanks to LJ Zigerell, Peter Frost, Noah Carl and Bryan Pesta for reviewing the paper and offering constructive criticism.

References

- Ashton, M. C., & Lee, K. (2005, July). Problems with the method of correlated vectors. *Intelligence*, 33(4), 431–444. doi: [10.1016/j.intell.2004.12.004](https://doi.org/10.1016/j.intell.2004.12.004)
- Beaver, K. M., & Wright, J. P. (2011, January). The association between county-level IQ and county-level crime rates. *Intelligence*, 39(1), 22–26. doi: [10.1016/j.intell.2010.12.002](https://doi.org/10.1016/j.intell.2010.12.002)
- Bogg, T., & Roberts, B. W. (2004). Conscientiousness and Health-Related Behaviors: A Meta-Analysis of the Leading Behavioral Contributors to Mortality. *Psychological Bulletin*, 130(6), 887–919. doi: [10.1037/0033-2909.130.6.887](https://doi.org/10.1037/0033-2909.130.6.887)
- Bogg, T., & Roberts, B. W. (2013, June). The Case for Conscientiousness: Evidence and Implications for a Personality Trait Marker of Health and Longevity. *Annals of Behavioral Medicine*, 45(3), 278–288. doi: [10.1007/s12160-012-9454-6](https://doi.org/10.1007/s12160-012-9454-6)
- Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015, January). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics*, 96(1), 37–53. doi: [10.1016/j.ajhg.2014.11.010](https://doi.org/10.1016/j.ajhg.2014.11.010)
- Costa Jr., P., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81(2), 322–331. doi: [10.1037/0022-3514.81.2.322](https://doi.org/10.1037/0022-3514.81.2.322)
- De Fruyt, F., & Mervielde, I. (1999, September). Riasec Types and Big Five Traits as Predictors of Employment Status and Nature of Employment. *Personnel Psychology*, 52(3), 701–727. doi: [10.1111/j.1744-6570.1999.tb00177.x](https://doi.org/10.1111/j.1744-6570.1999.tb00177.x)
- Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between mars and venus: measuring global sex differences in personality. *PLoS One*, 7(1), e29265. doi: <https://doi.org/10.1371/journal.pone.0029265>

- Deryugina, T., & Hsiang, S. (2014, December). *Does the Environment Still Matter? Daily Temperature and Income in the United States* (Tech. Rep. No. w20750). Cambridge, MA: National Bureau of Economic Research. Retrieved 2016-04-01, from <http://www.nber.org/papers/w20750.pdf>
- Diniz-Filho, J. A. F., Nabout, J. C., de Campos Telles, M. P., Soares, T. N., & Rangel, T. F. L. V. B. (2009). A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and Molecular Biology*, 32(2), 203–211. doi: [10.1590/S1415-47572009000200001](https://doi.org/10.1590/S1415-47572009000200001)
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A Meta-Analytic Investigation of Conscientiousness in the Prediction of Job Performance: Examining the Intercorrelations and the Incremental Validity of Narrow Traits. *Journal of Applied Psychology*, 91(1), 40–57. doi: [10.1037/0021-9010.91.1.40](https://doi.org/10.1037/0021-9010.91.1.40)
- Eagly, A. H. (2016, March). When Passionate Advocates Meet Research on Diversity, Does the Honest Broker Stand a Chance? *Journal of Social Issues*, 72(1), 199–222. doi: [10.1111/josi.12163](https://doi.org/10.1111/josi.12163)
- Everitt, B., & Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. New York, NY: Springer New York. Retrieved 2016-05-08, from <http://link.springer.com/10.1007/978-1-4419-9650-3>
- Fischer, D. H. (1989). *Albion's seed: Four British folkways in America*. Oxford University Press. Retrieved 2015-07-28, from <https://www.google.com/books?hl=en&lr=&id=eq5oAgAAQBAJ&oi=fnd&pg=PP1&dq=albion%27s+seed&ots=0-fe5Ym0ag&sig=cHk3eK-DjbOU3uP4Yew5TI0KSCY>
- Fuerst, J. (2014, jul). Ethnic/race differences in aptitude by generation in the united states: An exploratory meta-analysis. *Open Differential Psychology*. Retrieved from <https://doi.org/10.26775%2Fodp.2014.07.26> doi: [10.26775/odp.2014.07.26](https://doi.org/10.26775/odp.2014.07.26)
- Fuerst, J., & Kirkegaard, E. O. W. (2016a). Admixture in the americas: Regional and national differences. *Mankind Q.*, 56(3), 256–374. doi: [10.46469/mq.2016.56.3.2](https://doi.org/10.46469/mq.2016.56.3.2)
- Fuerst, J., & Kirkegaard, E. O. W. (2016b). The genealogy of differences in the americas. *Mankind Q.*, 56(3), 426–482. doi: [10.46469/mq.2016.56.3.9](https://doi.org/10.46469/mq.2016.56.3.9)
- Gelade, G. A. (2008, November). The geography of IQ. *Intelligence*, 36(6), 495–501. Retrieved 2015-09-06, from <http://www.sciencedirect.com/science/article/pii/S0160289608000202> doi: [10.1016/j.intell.2008.01.004](https://doi.org/10.1016/j.intell.2008.01.004)
- Giudice, M. D., Booth, T., & Irwing, P. (2012, January). The Distance Between Mars and Venus: Measuring Global Sex Differences in Personality. *PLOS ONE*, 7(1), e29265. doi: [10.1371/journal.pone.0029265](https://doi.org/10.1371/journal.pone.0029265)
- Hassall, C., & Sherratt, T. N. (2011). Statistical inference and spatial patterns in correlates of IQ. *Intelligence*, 39(5), 303–310. doi: [10.1016/j.intell.2011.05.001](https://doi.org/10.1016/j.intell.2011.05.001)
- Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis correcting error and bias in research findings*. Thousand Oaks, Calif.: Sage. Retrieved 2014-11-29, from <http://site.ebrary.com/id/10387875>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in r*. New York: Springer.
- Jensen, A. R. (1998). *The g factor: the science of mental ability*. Westport, Conn.: Praeger.
- Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36(1), 81–95. doi: <https://doi.org/10.1016/j.intell.2007.06.001>
- Kirkegaard, E. O. W. (2014a, oct). Crime, income, educational attainment and employment among immigrant groups in Norway and Finland. *Open Differential Psychology*. Retrieved from <https://doi.org/10.26775%2Fodp.2014.10.09> doi: [10.26775/odp.2014.10.09](https://doi.org/10.26775/odp.2014.10.09)
- Kirkegaard, E. O. W. (2014b, mar). Criminality and fertility among Danish immigrant populations. *Open Differential Psychology*. Retrieved from <https://doi.org/10.26775%2Fodp.2014.03.24> doi: [10.26775/odp.2014.03.24](https://doi.org/10.26775/odp.2014.03.24)

- Kirkegaard, E. O. W. (2014c, sep). The international general socioeconomic factor: Factor analyzing international rankings. *Open Differential Psychology*. Retrieved from <https://doi.org/10.26775%2Fodp.2014.09.08> doi: 10.26775/odp.2014.09.08
- Kirkegaard, E. O. W. (2015a). Examining the S factor in US states. Retrieved from <https://thewinnower.com/papers/examining-the-s-factor-in-us-states>
- Kirkegaard, E. O. W. (2015b, September 23). *The general religious factor among Muslims: a multi-level factor analysis*. Retrieved from <https://emilkirkegaard.dk/en/2015/09/the-general-religious-factor-among-muslims-a-multi-level-factor-analysis/>
- Kirkegaard, E. O. W. (2015c). A replication of the S factor among US states using a new and larger dataset. Retrieved from <https://thewinnower.com/papers/1418-a-replication-of-the-s-factor-among-us-states-using-a-new-and-larger-dataset>
- Kirkegaard, E. O. W. (2015d). Some methods for measuring and correcting for spatial autocorrelation. Retrieved from <https://thewinnower.com/papers/2847-some-methods-for-measuring-and-correcting-for-spatial-autocorrelation>
- Kirkegaard, E. O. W. (2016a, jan). Inequality among 32 london boroughs: An s factor analysis. *Open Quantitative Sociology & Political Science*. Retrieved from <https://doi.org/10.26775%2Foqsp.2016.01.26> doi: 10.26775/oqsp.2016.01.26
- Kirkegaard, E. O. W. (2016b, nov). Some new methods for exploratory factor analysis of socioeconomic data. *Open Quantitative Sociology & Political Science*. Retrieved from <https://doi.org/10.26775%2Foqsp.2016.11.07> doi: 10.26775/oqsp.2016.11.07
- Kirkegaard, E. O. W., & Fuerst, J. (2016). Inequality in the united states: Ethnicity, racial admixture and environmental causes. *Mankind Q.*, 56(4), 580–606. doi: 10.46469/mq.2016.56.4.7
- León, F. R. (2016). Race vis-à-vis latitude: Their influence on intelligence, infectious diseases, and income. *Mankind Q.*, 56(3), 412–420. doi: 10.46469/mq.2016.56.3.7
- León, F. R. (2015, November). The east-to-west decay of math and reading scores in the United States: A prediction from UVB radiation theory. *Personality and Individual Differences*, 86, 287–290. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0191886915004092> doi: 10.1016/j.paid.2015.06.028
- León, F. R., & Burga-León, A. (2015, May). How geography influences complex cognitive ability. *Intelligence*, 50, 221–227. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0160289615000586> doi: 10.1016/j.intell.2015.04.011
- Lippa, R. A. (2010, November). Gender Differences in Personality and Interests: When, Where, and Why? *Social and Personality Psychology Compass*, 4(11), 1098–1110. doi: 10.1111/j.1751-9004.2010.00320.x
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology (Gott.)*, 2(2), 57–64. doi: 10.1027/1614-2241.2.2.57
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "sinking shafts at a few critical points". *Annual Review of Psychology*, 51, 405–444. doi: 10.1146/annurev.psych.51.1.405
- Lumley, T. (2014, August). *survey: analysis of complex survey samples*. Retrieved 2016-04-02, from <https://cran.r-project.org/web/packages/survey/index.html>
- Major, J. T., Johnson, W., & Bouchard Jr., T. J. (2011, September). The dependability of the general factor of intelligence: Why small, single-factor models do not adequately represent g. *Intelligence*, 39(5), 418–433. Retrieved 2016-03-26, from <http://www.sciencedirect.com/science/article/pii/S0160289611000857> doi: 10.1016/j.intell.2011.07.002
- Meer, T. v. d., & Tolsma, J. (2014). Ethnic Diversity and Its Effects on Social Cohesion. *Annual Review of Sociology*, 40(1), 459–478. Retrieved 2016-03-23, from <http://dx.doi.org/10.1146/annurev-soc-071913-043309> doi: 10.1146/annurev-soc-071913-043309
- Oberski, D. (2015, November). *lavaan.survey: Complex Survey Structural Equation Modeling (SEM)*. Retrieved 2016-04-02, from <https://cran.r-project.org/web/packages/lavaan.survey/index.html>

- Pesta, B. J. (2016). Does IQ cause race differences in well-being? *Mankind Q.*, 56(3), 421–425. doi: [10.46469/mq.2016.56.3.8](https://doi.org/10.46469/mq.2016.56.3.8)
- Pesta, B. J., & Poznanski, P. J. (2014, September). Only in America: Cold Winters Theory, race, IQ and well-being. *Intelligence*, 46, 271–274. Retrieved 2014-10-13, from <http://www.sciencedirect.com/science/article/pii/S0160289614001019> doi: [10.1016/j.intell.2014.07.009](https://doi.org/10.1016/j.intell.2014.07.009)
- Putterman, L., & Weil, D. N. (2010, November). Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality. *The Quarterly Journal of Economics*, 125(4), 1627–1682. Retrieved from <http://qje.oxfordjournals.org/lookup/doi/10.1162/qjec.2010.125.4.1627> doi: [10.1162/qjec.2010.125.4.1627](https://doi.org/10.1162/qjec.2010.125.4.1627)
- Radil, S. M. (2011). *Spatializing social networks: making space for theory in spatial analysis*. Retrieved from <https://www.ideals.illinois.edu/handle/2142/26222> (University of Illinois at Urbana-Champaign)
- Revelle, W. (2015a, October). Personality research: an open and shared science. Novi Sad, Serbia.
- Revelle, W. (2015b, April). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Retrieved 2015-04-29, from <http://cran.r-project.org/web/packages/psych/index.html>
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007, December). The Power of Personality. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 2(4), 313–345. Retrieved 2016-04-28, from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4499872/> doi: [10.1111/j.1745-6916.2007.00047.x](https://doi.org/10.1111/j.1745-6916.2007.00047.x)
- Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., ... Barendse, M. (2015, November). *lavaan: Latent Variable Analysis*. Retrieved 2016-04-02, from <https://cran.r-project.org/web/packages/lavaan/index.html>
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001, June). Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: A Meta-Analysis. *Personnel Psychology*, 54(2), 297–330. Retrieved 2015-11-27, from <http://onlinelibrary.wiley.com/doi/10.1111/j.1744-6570.2001.tb00094.x/abstract> doi: [10.1111/j.1744-6570.2001.tb00094.x](https://doi.org/10.1111/j.1744-6570.2001.tb00094.x)
- Schmitt, N., Golubovich, J., & Leong, F. T. L. (2011, December). Impact of Measurement Invariance on Construct Correlations, Mean Differences, and Relations With External Correlates An Illustrative Example Using Big Five and RIASEC Measures. *Assessment*, 18(4), 412–427. Retrieved 2016-04-28, from <http://asm.sagepub.com/content/18/4/412> doi: [10.1177/1073191110373223](https://doi.org/10.1177/1073191110373223)
- Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135(6), 859–884. doi: [10.1037/a0017364](https://doi.org/10.1037/a0017364)
- Tate, B. W., & McDaniel, M. A. (2008). *Race Differences in Personality: An Evaluation of Moderators and Publication Bias*. Retrieved from http://www.people.vcu.edu/~mamcdani/Publications/AOM_as_of_8%201%202008.pdf
- Templ, M., Alfons, A., Kowarik, A., & Prantner, B. (2015, February). *VIM: Visualization and Imputation of Missing Values*. CRAN. Retrieved from <http://cran.r-project.org/web/packages/VIM/index.html>
- Templ, M., Kowarik, A., & Filzmoser, P. (2011, October). Iterative stepwise regression imputation using standard and robust methods. *Computational Statistics & Data Analysis*, 55(10), 2793–2806. Retrieved 2016-03-23, from <http://www.sciencedirect.com/science/article/pii/S0167947311001411> doi: [10.1016/j.csda.2011.04.012](https://doi.org/10.1016/j.csda.2011.04.012)
- Thorndike, R. L. (1987). Stability of factor loadings. *Personality and Individual Differences*, 8(4), 585–586. Retrieved 2016-03-26, from <http://www.sciencedirect.com/science/article/pii/0191886987902248> doi: [10.1016/0191-8869\(87\)90224-8](https://doi.org/10.1016/0191-8869(87)90224-8)
- Walker, T. L., & Tracey, T. J. G. (2012, February). Perceptions of occupational prestige: Differences between African American and White college students. *Journal of Vocational Behavior*, 80(1), 76–81. Retrieved 2016-04-28, from <http://www.sciencedirect.com/science/article/pii/S0001879111000959> doi: [10.1016/j.jvb.2011.06.003](https://doi.org/10.1016/j.jvb.2011.06.003)
- Woods, S. A., & Hampson, S. E. (2010, November). Predicting Adult Occupational Environments from Gender and Childhood Personality Traits. *The Journal of applied psychology*, 95(6), 1045–1057. Retrieved 2016-04-28, from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2964386/> doi: [10.1037/a0020600](https://doi.org/10.1037/a0020600)