₁   EDLD 653 Final Project

₂   Anisha Babu, Ian Shryock, Dillon Welindt, Futing Zou, Diana DeWald[1]

₃   [1] University of Oregon

₄   Author Note

6   Abstract

7   Big Five personality data were collected from ~114,000 individuals in the US between

8   April 2006 and August 2010. Participants were randomly assigned to question blocks of an

9   assessment titled 'International Personality Item Pool Big Five Factor Markers,' to indicate

10 individuals' Big Five traits (Openness, Conscientiousness, Extraversion, Agreeableness, and

11 Neuroticism). Additionally, demographic data were collected from participants on an

12 optional basis. Such data included questions about participants' gender, age, country, state

13 (if applicable), race, and education level. We used this demographic data in tandem with

14 trait data to run descriptive, correlational, and factor analyses, hypothesizing that there

15 would be variation in the number of ideal dimensions across states. We found this to be

16 the case, with personality traits demonstrating best fit in a range of 5 to 8 traits across 50

17 states.

18   *Keywords:* Personality traits, SAPA-Project Dataverse, Functional Programming

19   Word count: 1283

EDLD 653 Final Project

Introduction

Big Five

One of the most widely replicated findings within the field of personality psychology is the Big Five structure of personality. With roots in the 1800's, personality psychology sought to determine the best way to represent the large number of personality traits in a concise structure. This research initially involved researchers providing participants with large numbers of trait descriptive adjectives and asking them to rate the extent to which those adjectives characterize themselves or someone they knew. Dimension reduction analyses were then used to create a simpler structure from those responses.

Multiple research groups began converging on the five factor structure as early as the 1960's, with an increasing consensus by the late 1980's. Most of the recent work on the big five has been conducted through a combination of confirmatory factor analysis and theory driven selection of survey items based on previous findings about the structure.

Geographical Personality

In recent years, there has been increasing focus on regional variation of personality traits within the United States. Work has examined the extent to which regions of the US differ on the Big Five domains and can be said to have distinct and characteristic combinations of trait levels. For example, Rentfrow and colleagues (2013) show that the south and midwest are best characterized as friendly and conventional, whereas the west is relaxed and creative, and the northeast is temperamental and uninhibited.

A limitation of this work is that it examines the extent to which the five factor structure captures each region and what differences in the levels of each factor are due to

43 regional variation. This research utilizes confirmatory factor analyses that assume that the
44 five factor structure is the ideal level of dimensionality to characterize all regions.

Cross-Cultural Studies

46 Much of the cross-cultural work on personality structure has found some support for
47 the notion that the five factor structure has applicability in a number of cultures. However,
48 these studies typically are conducted from an etic perspective that translate the items used
49 in western samples.

50 However, when studies are conducted from an emic perspective – that is, using trait
51 descriptive adjectives from the language of the culture, rather than translations of items
52 used in the big five framework – different structures emerge. A varying number of factors
53 have been found to best fit different cultures, ranging from one to seven in many cases.

Geographical Factor Structure within US

55 Within the US, the regional variation in factor structures has not been an extensively
56 studied topic. Because most research operates within a framework that utilizes
57 confirmatory factor analysis, there is little information on the extent to which regions differ
58 in their factor structure.

59 In the current study, we use exploratory factor analyses to provide estimates of the
60 optimal factor structures for each of the fifty states.

Methods

Measures

63 The International Personality Item Pool is an open-source repository of personality
64 trait items that have been researched extensively in the big five tradition. The current

study uses ninety nine of one hundred items from the IPIP-100. Participants rated themselves on a number of personality traits from 1- not at all like me to 6- very much like me.

Data Collection

Data were obtained from the Harvard Dataverse (D. Condon, Zabelina, and Revelle (2021)). Data were initially collected using the Synthetic Aperture for Personality Assessment (D. M. Condon & Revelle, 2014; see Revelle et al., 2016; Wilt, Funkhouser, & Revelle, 2011) which utilizes a massively missing completely at random design, wherein each participant only provides responses to a fraction of items.

Data analysis

First, we provide descriptive norms for the entire US sample, and then by state.

Next, we use parallel analysis to determine the optimal number of factors in the whole sample. Our hypothesis is that five factors will provide an optimal fit.

The main analyses are fifty parallel analyses, one for every state, that estimates the optimal number of personality dimensions for each state. We hypothesize that there will be variation in the number of ideal dimensions across states.

Results

The parallel factor analyses indicate that the modal number of factors is 5, as is

found in 36 of the 50 states, with 7, 6, and 1 states respectively being better represented by

6, 7, and 8 factors. This analysis relies on covariance matrices uing pairwise complete

observations; use of a raw matrix does not converge for each state. However, where such

analyses are convergent, findings indicate a factor structure of approximately 20-30.

Personality Structure for United States

Personality Structure for Florida

88
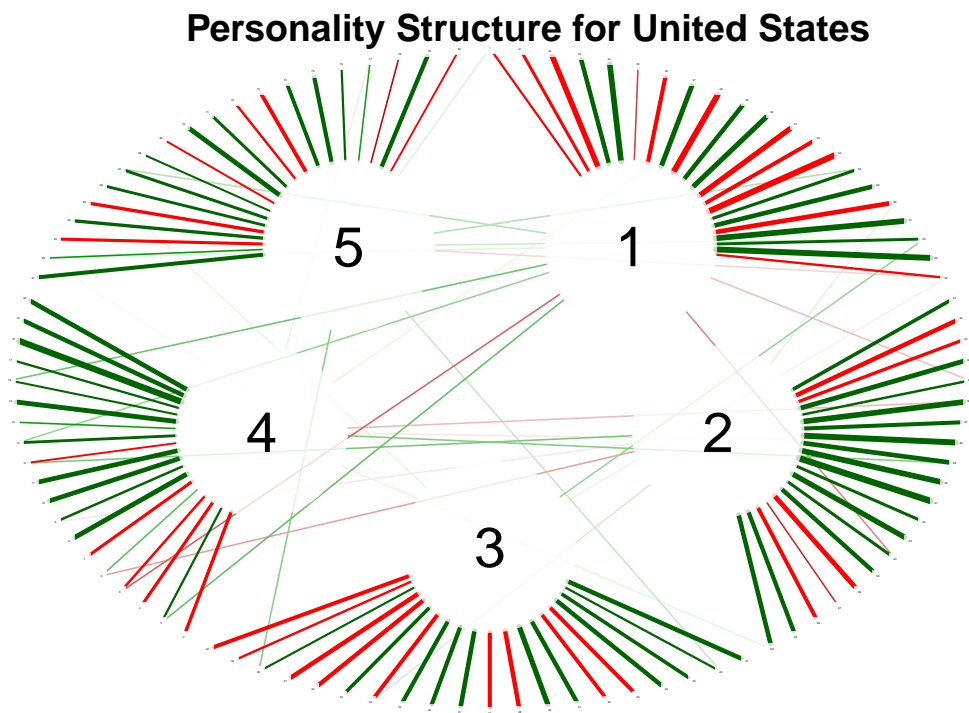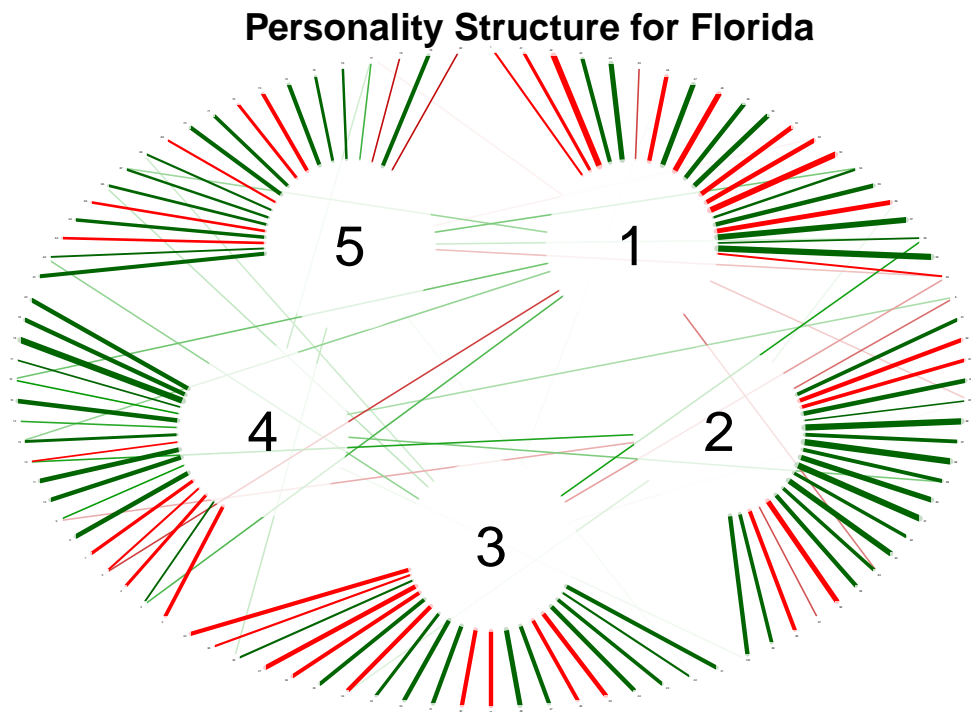
Personality Structure for Arkansas

89
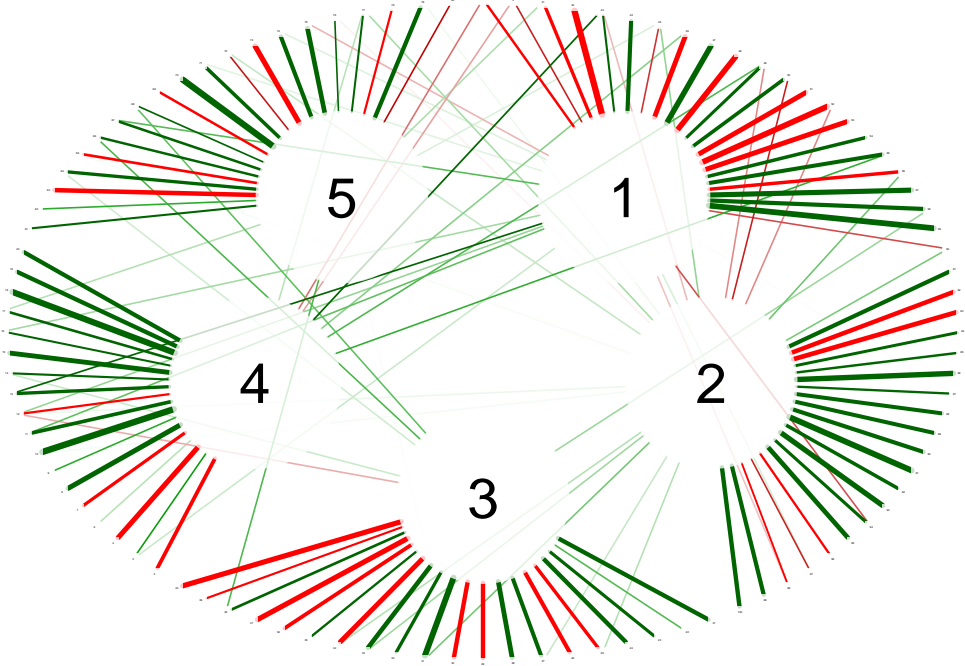
Personality Structure for North Dakota

90

Discussion

92 These findings appear to confirm the Big Five factor structure of personality.

93 However, this conclusion is extremely tentative, and relies on numerous assumptions

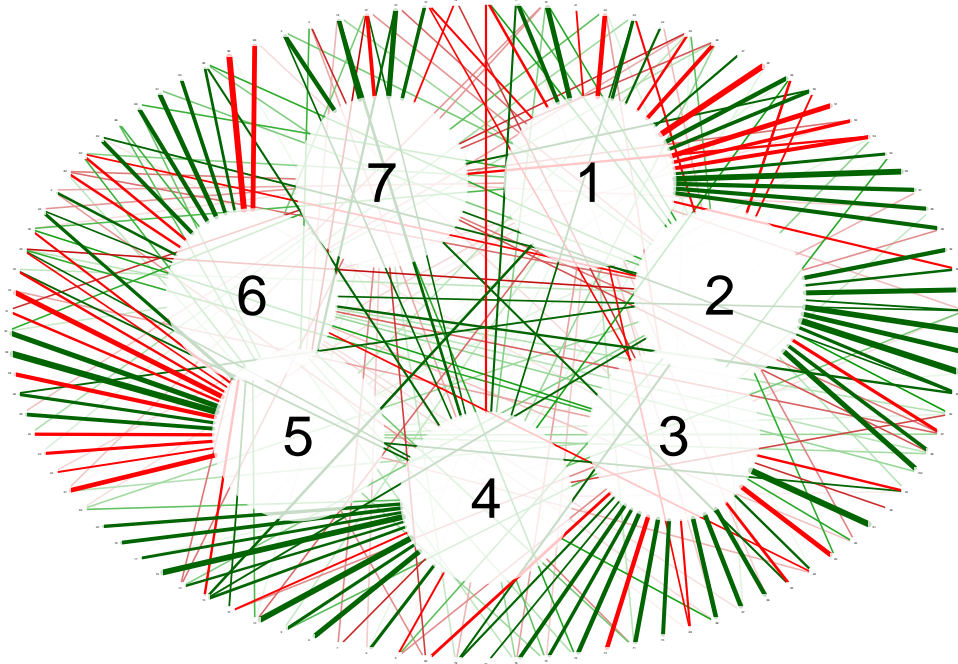94 holding, for which there is little basis. Most notably, the Kaiser, Meyer, Olkin Measure of

95 Sampling Adequacy indicates a "Marvelous" (Kaiser (1974)) measure of the whole sample,

96 whereas no state on it's own is acceptable for factor analysis by those same standards. The

97 complete sample appears to have a 20-factor structure which is more consistent with the

98 findings of related SAPA studies, which finds a 27-factor structure. That the complete

99 sample yielded a 20-factor structure in our exploratory analysis, suggests: 1. an issue of

100 completeness limiting state-level inference (an expected outcome of the SAPA sampling

101 strategy), and; 2. A potential hierarchical factor structure in which the 20 factors

102 aggregate to fewer.

¹⁰³ However, barring this issue, it appears notable that smaller states are better
¹⁰⁴ represented by more than five factors. This is surprising, as there were no a priori data of
¹⁰⁵ the populations of these states entered into the analyses, and thus, such inferences are not
¹⁰⁶ based on a more robust estimate. Further inspection demonstrates that states with a larger
¹⁰⁷ number of factors have a number of responses approximately commensurate with their state
¹⁰⁸ population size; that is to say, less populous states produce fewer observations and a higher
¹⁰⁹ factor structure. It seems a reasonable hypothesis that these findings are not well-founded.

¹¹⁰ Thus, while state-level (dis)confirmation of any extant published personality
¹¹¹ structure, and demonstration of the homogeneity therein remains a fruitful endeavor, these
¹¹² data do not admit to being interpreted to those ends. SAPA, relying on large degrees of
¹¹³ missingness, is not well-suited for such a degree of granularity. A shorter, but more
¹¹⁴ complete personality questionnaire would be more apt. However, stable measures of an
¹¹⁵ individual's personality is a function of the number of items, and thus limits the number of
¹¹⁶ factors to extract. Thus, there is a tension between the so-called "Big Few" and personality
¹¹⁷ structures comprised of more numerous factors. Due to the limited attention of any
¹¹⁸ respondent, and limited resources of researchers to incentivize larger questionnaires, the
¹¹⁹ Big Few personality structure will continue to be reified; if factors aggregate to form this
¹²⁰ structure, this measurement is limited but valid. To the extent that they do not, these
¹²¹ prevailing structures are increasingly problematic.

122   We used the following R packages for this manuscript: R [Version 4.1.1; R Core Team

123 (2021)] and the R-packages *apaTables* [Version 2.0.8; Stanley (2021)], *arsenal* [Version

124 3.6.3; Heinzen, Sinnwell, Atkinson, Gunderson, and Dougherty (2021)], *dataverse* [Version

125 0.3.10; Kuriwaki, Beasley, and Leeper (2022)], *dplyr* [Version 1.0.9; Wickham, François,

126 Henry, and Müller (2022)], *forcats* [Version 0.5.1; Wickham (2021)], *ggplot2* [Version 3.3.6;

127 Wickham (2016)], *GPArotation* [Version 2022.4.1; Bernaards and I.Jennrich (2005)], *here*

128 [Version 1.0.1; Müller (2020)], *kableExtra* [Version 1.3.4; Zhu (2021)], *needs* [Version 0.0.3;

129 Katz (2016)], *papaja* [Version 0.1.0.9997; Aust and Barth (2020)], *psych* [Version 2.2.5;

130 Revelle (2022)], *purrr* [Version 0.3.4; Henry and Wickham (2020)], *qgraph* [Version 1.9.2;

131 Epskamp, Cramer, Waldorp, Schmittmann, and Borsboom (2012)], *readr* [Version 2.1.2;

132 Wickham, Hester, and Bryan (2022)], *rio* [Version 0.5.29; Chan, Chan, Leeper, and Becker

133 (2021)], *stringr* [Version 1.4.0; Wickham (2019)], *tibble* [Version 3.1.7; Müller and Wickham

134 (2022)], *tidyr* [Version 1.2.0; Wickham and Girlich (2022)], and *tidyverse* [Version 1.3.1;

135 Wickham et al. (2019)]

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Bernaards, C. A., & I.Jennrich, R. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*, 676–696.

Chan, C., Chan, G. C., Leeper, T. J., & Becker, J. (2021). *Rio: A swiss-army knife for data file i/o.*

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64.

Condon, D., Zabelina, D., & Revelle, W. (2021). *Reproducibility Data for: Creative Achievement and Individual Differences* (Version V3) [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/2IBBMG

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18.

Heinzen, E., Sinnwell, J., Atkinson, E., Gunderson, T., & Dougherty, G. (2021). *Arsenal: An arsenal of 'r' functions for large-scale statistical summaries.* Retrieved from https://CRAN.R-project.org/package=arsenal

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36.

Katz, J. (2016). *Needs: Attaches and installs packages.* Retrieved from https://CRAN.R-project.org/package=needs

Kuriwaki, S., Beasley, W., & Leeper, T. J. (2022). *Dataverse: R client for dataverse 4+ repositories.*

163   Müller, K. (2020). *Here: A simpler way to find your files.* Retrieved from

164   https://CRAN.R-project.org/package=here

165   Müller, K., & Wickham, H. (2022). *Tibble: Simple data frames.* Retrieved from

166   https://CRAN.R-project.org/package=tibble

167   R Core Team. (2021). *R: A language and environment for statistical computing.*

168   Vienna, Austria: R Foundation for Statistical Computing. Retrieved from

169   https://www.R-project.org/

170   Revelle, W. (2022). *Psych: Procedures for psychological, psychometric, and*

171   *personality research.* Evanston, Illinois: Northwestern University. Retrieved from

172   https://CRAN.R-project.org/package=psych

173   Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G.

174   (2016). Web and phone based data collection using planned missing designs.

175   *Sage Handbook of Online Research Methods (2nd Ed., P. 578-595). Sage*

176   *Publications, Inc.*

177   Stanley, D. (2021). *apaTables: Create american psychological association (APA)*

178   *style tables.* Retrieved from https://CRAN.R-project.org/package=apaTables

179   Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag

180   New York. Retrieved from https://ggplot2.tidyverse.org

181   Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string*

182   *operations.* Retrieved from https://CRAN.R-project.org/package=stringr

183   Wickham, H. (2021). *Forcats: Tools for working with categorical variables (factors).*

184   Retrieved from https://CRAN.R-project.org/package=forcats

185   Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . .

186   Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software,*

187   *4*(43), 1686. https://doi.org/10.21105/joss.01686

188   Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of*

189   *data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

190   Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data.* Retrieved from

191   https://CRAN.R-project.org/package=tidyr

192   Wickham, H., Hester, J., & Bryan, J. (2022). *Readr: Read rectangular text data.*

193   Retrieved from https://CRAN.R-project.org/package=readr

194   Wilt, J., Funkhouser, K., & Revelle, W. (2011). The dynamic relationships of

195   affective synchrony to perceptions of situations. *Journal of Research in*

196   *Personality, 45*(3), 309–321.

197   Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax.*

198   Retrieved from https://CRAN.R-project.org/package=kableExtra

Table 1

Race Demographics

Race	N	Prop. of Sample
African American	6108	0.0750682
Chinese	1129	0.0138756
Indian/Pakistani	469	0.0057641
Japanese	257	0.0031586
Korean	500	0.0061451
Latino	2079	0.0255512
Mexican	2166	0.0266205
Native American	728	0.0089472
Other	3067	0.0376939
Other Asian	566	0.0069562
Pacific Islander	305	0.0037485
Philipino	615	0.0075584
Puerto Rican	512	0.0062926
White/Caucasian	62859	0.7725463
NA	6	0.0000737

Table 2

Education Demographics

Education	N	Prop.of Sample
College graduate	12381	0.1521643
Currently attending college	32469	0.3990487
Graduate or professional degree	10338	0.1270555
High school graduate	6145	0.0755229
Less than 12 years	11759	0.1445198
Some college did not graduate	8274	0.1016887

Table 3

Agreeableness Descriptives

State	Mean	SD	N
Alabama	4.610537	0.7744197	643
Alaska	4.629332	0.7841253	555
Arizona	4.596157	0.7729134	866
Arkansas	4.689144	0.7750065	577
California	4.663943	0.7623445	9709
Colorado	4.631178	0.7809921	1097
Connecticut	4.634948	0.7713774	986
Delaware	4.575882	0.7144243	592
Florida	4.659417	0.8070510	2936
Georgia	4.688953	0.7696673	2414
Hawaii	4.637072	0.8284047	292
Idaho	4.672925	0.7444675	340
Illinois	4.729347	0.7352130	5520
Indiana	4.689507	0.7638457	1707
Iowa	4.667843	0.7655999	982
Kansas	4.678785	0.7617303	808
Kentucky	4.693537	0.7580447	820
Louisiana	4.703087	0.7277572	2030
Maine	4.667634	0.7913389	356
Maryland	4.707332	0.7694775	1772
Massachusetts	4.654854	0.7788784	1935
Michigan	4.653881	0.7815882	2549

Table 3

Agreeableness Descriptives (continued)

State	Mean	SD	N
Minnesota	4.667296	0.7423893	2104
Mississippi	4.722192	0.7991229	604
Missouri	4.665174	0.7644602	1611
Montana	4.765261	0.6641912	243
Nebraska	4.641667	0.7724856	580
Nevada	4.633536	0.7698699	274
New Hampshire	4.629999	0.7428263	389
New Jersey	4.669726	0.7642500	2495
New Mexico	4.727511	0.7475549	1199
New York	4.663282	0.8054766	4942
North Carolina	4.647067	0.7780090	1454
North Dakota	4.615877	0.8189782	190
Ohio	4.700273	0.7612682	3600
Oklahoma	4.652295	0.7867416	771
Oregon	4.633837	0.7672260	1203
Pennsylvania	4.656941	0.7579218	4758
Rhode Island	4.732280	0.7377559	422
South Carolina	4.727500	0.7352295	1010
South Dakota	4.657946	0.8397148	172
Tennessee	4.663913	0.8098826	1133
Texas	4.652647	0.7789094	4662
Utah	4.665389	0.7546368	487

Table 3

Agreeableness Descriptives (continued)

State	Mean	SD	N
Vermont	4.776898	0.6847462	161
Virginia	4.722891	0.7433986	2787
Washington	4.653790	0.7633607	1742
West Virginia	4.618171	0.8055191	384
Wisconsin	4.641089	0.7689811	2377
Wyoming	4.662522	0.7447127	126

Table 4

Conscientiousness Descriptives

State	Mean	SD	N
Alabama	4.117571	0.9667964	643
Alaska	4.026173	0.8683283	555
Arizona	4.098011	0.8929967	866
Arkansas	4.147743	0.9248194	577
California	4.095768	0.9086644	9709
Colorado	4.124273	0.9210228	1097
Connecticut	4.120340	0.9307941	986
Delaware	3.997203	0.8632940	592
Florida	4.179669	0.9343133	2936
Georgia	4.129606	0.9180854	2414
Hawaii	4.170548	0.8806935	292
Idaho	4.150588	0.8735417	340
Illinois	4.162300	0.9112091	5520
Indiana	4.218357	0.9211999	1707
Iowa	4.102082	0.9230007	982
Kansas	4.108794	0.9105627	808
Kentucky	4.122991	0.9537198	820
Louisiana	4.209949	0.8955529	2030
Maine	4.205641	0.9188266	356
Maryland	4.128232	0.9095782	1772
Massachusetts	4.118530	0.9365527	1935
Michigan	4.162636	0.9299835	2549

Table 4

Conscientiousness Descriptives (continued)

State	Mean	SD	N
Minnesota	4.096184	0.8982866	2104
Mississippi	4.198448	0.9059218	604
Missouri	4.141600	0.9351437	1611
Montana	4.160722	0.9577293	243
Nebraska	4.156379	0.8777856	580
Nevada	4.082401	0.9495511	274
New Hampshire	4.148700	0.8946446	389
New Jersey	4.148890	0.9280449	2495
New Mexico	4.262791	0.8985910	1199
New York	4.148340	0.9339365	4942
North Carolina	4.171695	0.9481568	1454
North Dakota	4.287222	0.8787829	190
Ohio	4.188897	0.9247682	3600
Oklahoma	4.140532	0.9454096	771
Oregon	4.070511	0.9039862	1203
Pennsylvania	4.117532	0.9191805	4758
Rhode Island	4.099572	0.9192980	422
South Carolina	4.136447	0.8851777	1010
South Dakota	4.266537	0.8999202	172
Tennessee	4.208465	0.9470666	1133
Texas	4.131076	0.9285345	4662
Utah	4.102647	0.8660975	487

Table 4

Conscientiousness Descriptives (continued)

State	Mean	SD	N
Vermont	4.090649	0.9496127	161
Virginia	4.141444	0.9057514	2787
Washington	4.127309	0.9222153	1742
West Virginia	4.163824	0.9387757	384
Wisconsin	4.112204	0.9268177	2377
Wyoming	4.188095	0.9480086	126

Table 5

Extraversion Descriptives

State	Mean	SD	N
Alabama	3.744427	1.0737088	643
Alaska	3.802078	1.0298194	555
Arizona	3.826485	1.0391321	866
Arkansas	3.826526	1.1045412	577
California	3.916162	1.0182084	9709
Colorado	3.811444	1.0267798	1097
Connecticut	3.924073	1.0262157	986
Delaware	4.029242	0.9524545	592
Florida	3.888615	1.0586144	2936
Georgia	3.998944	1.0429965	2414
Hawaii	3.835455	1.0309916	292
Idaho	3.752173	1.0263106	340
Illinois	4.012087	0.9837590	5520
Indiana	3.894970	1.0577104	1707
Iowa	3.907930	1.0020006	982
Kansas	3.939745	1.0499232	808
Kentucky	3.938068	1.0360986	820
Louisiana	4.010502	0.9748960	2030
Maine	3.846177	1.0208230	356
Maryland	3.919828	1.0085951	1772
Massachusetts	3.894588	1.0272060	1935
Michigan	3.896950	1.0384003	2549

Table 5

Extraversion Descriptives (continued)

State	Mean	SD	N
Minnesota	3.941308	1.0015734	2104
Mississippi	3.920760	1.0388218	604
Missouri	3.912232	0.9806955	1611
Montana	3.847828	1.0609354	243
Nebraska	3.918702	0.9855573	580
Nevada	3.816920	1.0112589	274
New Hampshire	3.862380	0.9731819	389
New Jersey	3.996112	0.9864916	2495
New Mexico	3.898791	1.0508917	1199
New York	3.933406	1.0287508	4942
North Carolina	3.802822	1.0565457	1454
North Dakota	3.803845	1.0412612	190
Ohio	3.920517	1.0329506	3600
Oklahoma	3.804182	1.0792374	771
Oregon	3.886437	1.0061902	1203
Pennsylvania	3.958699	1.0067223	4758
Rhode Island	4.057464	0.9397830	422
South Carolina	4.043584	0.9876283	1010
South Dakota	3.979409	1.0081303	172
Tennessee	3.847732	1.0370815	1133
Texas	3.888641	1.0537326	4662
Utah	3.884845	1.0514536	487

Table 5

Extraversion Descriptives (continued)

State	Mean	SD	N
Vermont	3.917118	0.9878257	161
Virginia	3.943700	1.0111944	2787
Washington	3.809801	1.0334914	1742
West Virginia	3.779065	1.0798027	384
Wisconsin	3.920540	1.0092239	2377
Wyoming	3.821847	1.0187586	126

Table 6

Intellect Descriptives

State	Mean	SD	N
Alabama	4.637855	0.7449558	643
Alaska	4.664919	0.7668085	555
Arizona	4.665397	0.7510410	866
Arkansas	4.601727	0.7666522	577
California	4.607159	0.7273981	9709
Colorado	4.670469	0.7275069	1097
Connecticut	4.655686	0.7412064	986
Delaware	4.386655	0.7205952	592
Florida	4.657286	0.7032552	2936
Georgia	4.597719	0.7234032	2414
Hawaii	4.533509	0.7517126	292
Idaho	4.676192	0.7004048	340
Illinois	4.568339	0.7243646	5520
Indiana	4.578118	0.7459488	1707
Iowa	4.533963	0.7319487	982
Kansas	4.600704	0.7605258	808
Kentucky	4.601366	0.7377464	820
Louisiana	4.421264	0.7414032	2030
Maine	4.643924	0.7361798	356
Maryland	4.577738	0.7187010	1772
Massachusetts	4.574818	0.7139949	1935
Michigan	4.656310	0.7285229	2549

Table 6

Intellect Descriptives (continued)

State	Mean	SD	N
Minnesota	4.531498	0.7231444	2104
Mississippi	4.559547	0.7368477	604
Missouri	4.588972	0.7330890	1611
Montana	4.708861	0.7378293	243
Nebraska	4.495270	0.7481413	580
Nevada	4.651490	0.7185546	274
New Hampshire	4.630421	0.7565377	389
New Jersey	4.612689	0.7390479	2495
New Mexico	4.579827	0.6900077	1199
New York	4.649041	0.7256438	4942
North Carolina	4.569312	0.7502207	1454
North Dakota	4.570237	0.7177782	190
Ohio	4.556529	0.7364082	3600
Oklahoma	4.605553	0.7834762	771
Oregon	4.619854	0.7436931	1203
Pennsylvania	4.512731	0.7411906	4758
Rhode Island	4.622255	0.6808287	422
South Carolina	4.485802	0.7222830	1010
South Dakota	4.653013	0.6828377	172
Tennessee	4.590234	0.7736854	1133
Texas	4.613907	0.7484548	4662
Utah	4.607547	0.7199137	487

Table 6

Intellect Descriptives (continued)

State	Mean	SD	N
Vermont	4.735498	0.7363306	161
Virginia	4.547663	0.7079233	2787
Washington	4.684024	0.6893191	1742
West Virginia	4.582448	0.7334187	384
Wisconsin	4.504417	0.7397412	2377
Wyoming	4.631488	0.6969593	126