

Short Communication

Problems with the Big Five assessment in the World Values Survey

Steven G. Ludeke^{a,b,*}, Erik Gahner Larsen^{b,1}^a Department of Psychology, University of Southern Denmark, Odense, Denmark^b Department of Political Science, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Article history:

Received 30 December 2016

Received in revised form 16 February 2017

Accepted 18 February 2017

Keywords:

Big Five

World Values Survey

Reliability

BFI-10

ABSTRACT

Publicly-available data from the World Values Survey (WVS) is an extremely valuable resource for social scientists, serving as the basis for thousands of research publications. The most recent assessment (Wave 6) was the first to assess Big Five personality traits, and this data has already been used in published research. In the present paper, we show for the first time that the Big Five data from WVS Wave 6 is extremely problematic: items from the same trait correlate negatively with each other as often as not, occasionally to truly extreme degrees. Particular caution is warranted for any future research aiming to use this data, as we do not identify any straightforward solution to the data's challenges.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The six assessment waves collected thus far by the World Values Survey (WVS) represent one of the most heavily utilized empirical resources in all of social science, with the WVS organization claiming that thousands of publications are based on this publicly available data (World Values Survey, n.d.). The most recent assessment (Wave 6) included a validated short-form measure of the Big Five, the 10 item Big Five Inventory (BFI-10; Rammstedt & John, 2007). The BFI-10 has demonstrated admirable test-retest reliability despite its brevity, and previous cross-cultural studies using the long-form of the instrument has indicated these traits replicate reasonably well across cultural contexts (Rammstedt & John, 2007; Schmitt, Allik, McCrae, & Benet-Martinez, 2007). Of course, the assessment of personality traits outside of “WEIRD” (Western, Educated, Industrialized, Rich, and Democratic) contexts brings some challenges, such as lower levels of internal consistency reliability in responses in non-WEIRD samples (Church, 2001; Schmitt et al., 2007). However, the BFI-10 might be a good choice to attempt to remedy this deficit given that previous research on WEIRD samples has suggested BFI-10 items exhibit uncommonly high within-trait correlations, presumably reflecting the centrality of these items for their respective traits (Ziegler, Poropat, & Mell, 2014).

Given the richness of the World Values Survey data – with tens of thousands of participants in representative samples across more than 20 countries – it is unsurprising that published research (Fatke, 2016; Hanel & Vione, 2016) has already used this data despite its very recent release. However, this previous published research does not adequately evaluate basic considerations of the BFI-10 data from the WVS, and to our knowledge no public assessment has yet to do so.

As we show in Section 3, such considerations are particularly necessary for the data in question, as from a psychometric perspective they appear highly anomalous. Specifically, the inter-item correlations from each of the Big Five traits are clustered around 0, with negative values observed roughly as frequently as positive values. That is, two items from the same trait typically exhibited no relationship with each other. Some cross-country differences are consistent with previous research, with WEIRD nations having exhibited more typical patterns of inter-item correlations than did non-WEIRD nations.

2. Materials and methods

2.1. World Values Survey data collection

The World Values Survey has conducted face-to-face, nationally representative surveys in a multitude of different countries since 1981 and is the largest non-commercial survey in the world. We analyze data from Wave 6, collected between 2010 and 2014, which was the first to collect Big Five personality data. 59 countries participated in Wave 6. However, there is national variation in the inclusion of questions, and 25 countries collected data on the Big Five personality traits. We

* Corresponding author at: Department of Psychology, University of Southern, Campusvej 55, DK-5230 Odense, Denmark.

E-mail address: lude0011@umn.edu (S.G. Ludeke).

¹ Denotes equally contributing co-authors. Authorship order determined by rock-paper-scissors.

obtained the data from <http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>.

2.2. Measures

Participants completed the 10-item BFI, which includes two items per trait, one of which is reverse-coded. Valid responses were provided on a five-point Likert-type scale ranging from “Disagree strongly” to “Agree strongly,” with “neither agree nor disagree” as the midpoint. Responses of “don’t know” are treated as missing data. The R script required to reproduce the results presented here is provided in the online Supplementary materials.

2.3. Participants

The World Values Survey is typically limited to residents in the countries between the age of 18 and 85, with a few countries using 16 as the lower threshold. We focus on the respondents providing answers to all ten BFI-10 items. This results in a sample of 32,880 respondents (49.4% male) from 25 countries, with an average age of 40.5 (S.D. = 15.6). Ns for individual countries ranged between 653 (Yemen) and 3317 (South Africa).

3. Results

Correlation analyses were conducted separately for each country using the syntax provided in Supplementary materials. Full correlation

matrices for all BFI-10 items are presented there as well. The results of primary concern are presented in Fig. 1, which represents the within-trait inter-item correlations for each of the Big Five, after adjusting the reverse-coded item so that higher values should indicate higher levels of the trait.

Two things are immediately apparent from this figure. First, there is substantial variability in the inter-item correlations across countries. In particular, Germany and the Netherlands (the two WEIRD countries assessed) present results largely consistent with the intent of the instrument: all within-trait correlations were positive, and often of a magnitude approaching that expected based on previous studies using the BFI-10 ($r_s = .3$ to $.6$; see Ziegler et al., 2014). By contrast, all other countries exhibited markedly different results, with most even exhibiting negative inter-item correlations for at least one trait.

A second result of interest is that for many traits the typical inter-item correlation is abysmal. This can be seen even more readily in Fig. 2, which uses the same correlation output as Fig. 1. Here we observed three results of note. First, the inter-item correlations are distributed around zero. That is, two items intended to indicate the same trait typically exhibit no relationship whatsoever with each other. Second, the inter-item correlations are not equally problematic for all traits: inter-item correlations for Openness in fact tend to be negative, whereas those for Conscientiousness tend to be positive, if still more modest than that suggested by previous research (Ziegler et al., 2014). Extraversion, Agreeableness, and to a somewhat lesser extent Emotional Stability tend to exhibit little relationship between the two items. A third result of note is that responses in the Bahrain sample appear unique.

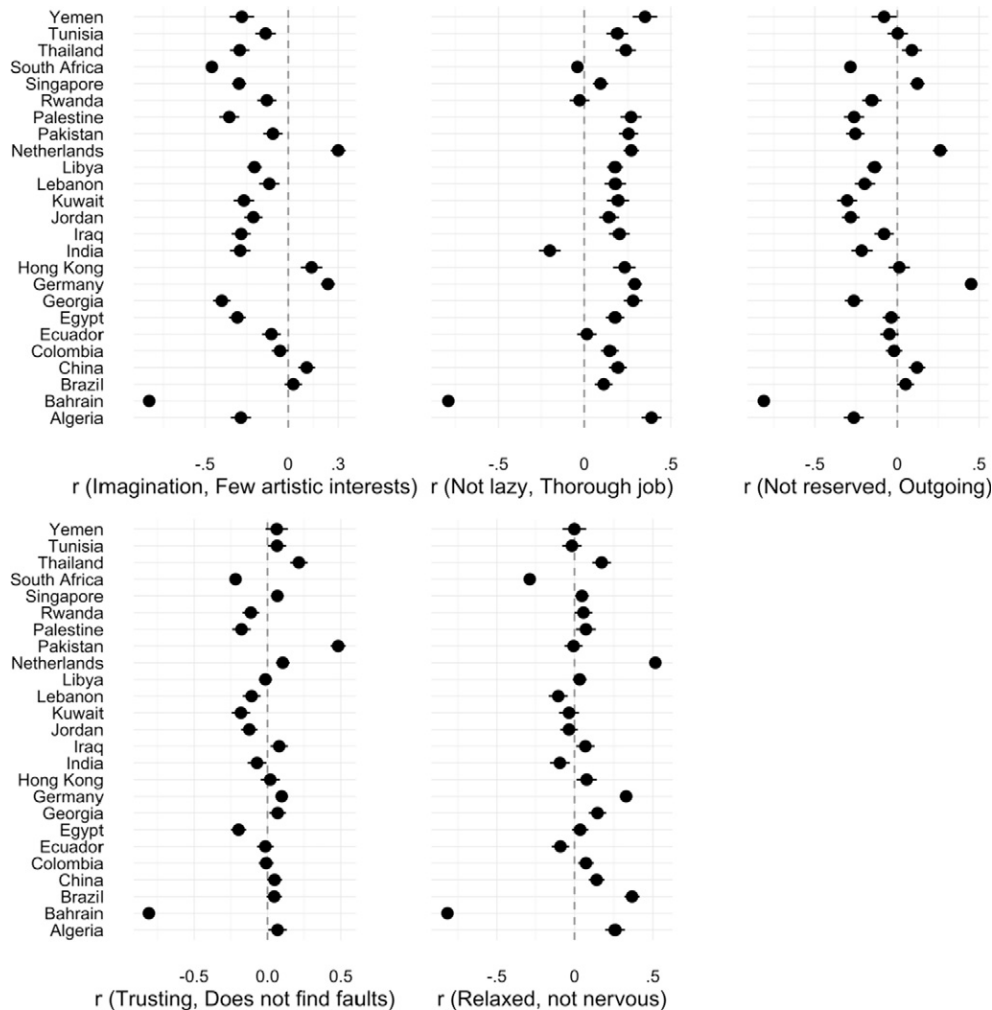


Fig. 1. Within-trait inter-item correlations for the BFI-10 by country. 95% confidence intervals provided for all countries. Countries differed in number of participants.

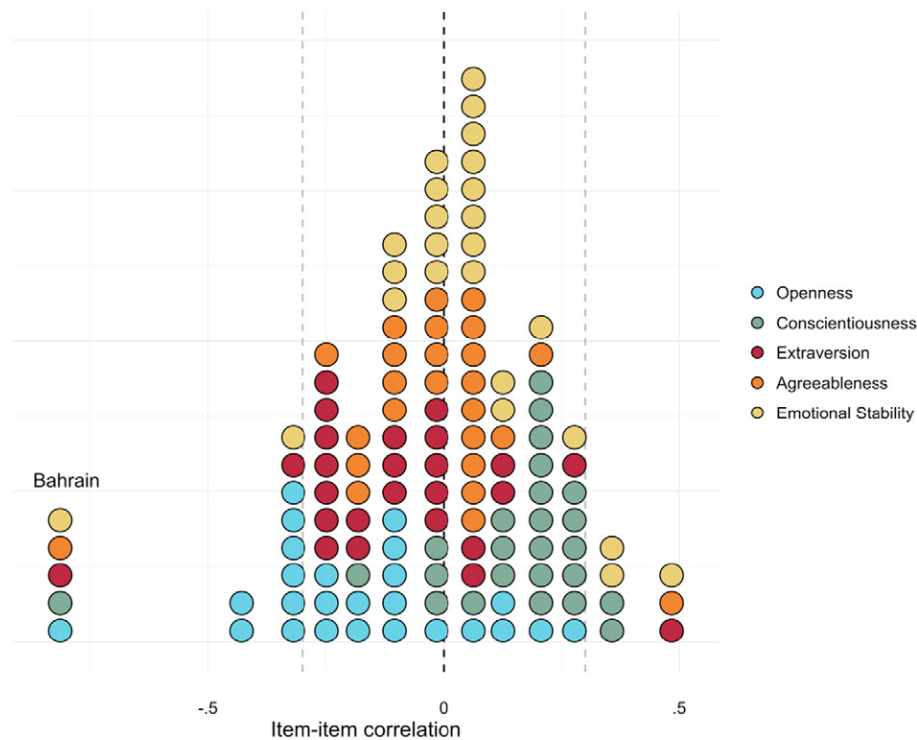


Fig. 2. Distribution of inter-item correlations in WVS Wave 6. Each dot represents the within-trait inter-item correlation obtained in one of the 25 countries with BFI data in WVS Wave 6. Gray dashed lines indicate correlations of $-.30$ and $.30$.

For all five traits, inter-item correlations are approximately $-.90$. In fact, as shown in Supplementary materials E, correlations between all BFI-10 items range between $|.80|$ and $|.90|$ in the Bahraini sample.

4. Discussion

In this brief communication we highlight a significant limitation to the BFI-10 data collected in Wave 6 of the World Values Survey. Although inter-item correlations in WEIRD countries were largely as-expected, overall there was typically no relationship between two items intended to assess a given trait.

Unfortunately, we do not see a straightforward explanation for these results. One might hypothesize that an error in data entry affected some countries but not others, such that the erroneously entered countries exhibited negative within-trait inter-item correlations. However, the inconsistent results across different traits within the same country speak against such an interpretation: nearly all countries exhibited counter-expectation results for Openness, but expectation-matching results for Conscientiousness. Furthermore, most countries exhibited minimal within-trait correlations between indicators of Extraversion, Emotional Stability, and especially of Agreeableness, suggesting a simple issue with regard to reverse-coded indicators cannot be the only challenge with this data. What's more, the simultaneous presence of (a) seemingly valid data from Germany and the Netherlands and (b) exceptionally strange data from Bahrain suggests the possible presence of country-specific errors as well.

Barring new understanding of how these curious results came about, we suggest it is thus hard to justify the use of this data in future research. If two items intended as the sole indicators of a trait do not correlate

with each other, there is no reason to suppose that either item provides any meaningful information about that trait in that population. Conclusions from existing published work using this data (e.g., Fatke, 2016; Hanel & Vione, 2016) should be carefully reconsidered in this light.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.paid.2017.02.042>.

References

- Church, A. T. (2001). Personality measurement in cross-cultural perspective. *Journal of Personality*, 69(6), 979–1006. <http://dx.doi.org/10.1111/1467-6494.696172>.
- Fatke, M. (2016). *Personality traits and political ideology: A first global assessment*. *Political Psychology*. <http://dx.doi.org/10.1111/pops.12347>.
- Hanel, P. H. P., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PloS One*, 11(12), e0168354. <http://dx.doi.org/10.1371/JOURNAL.PONE.0168354>.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212. <http://dx.doi.org/10.1016/j.jrp.2006.02.001>.
- Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martinez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 38(2), 173–212. <http://dx.doi.org/10.1177/0022022106297299>.
- World Values Survey. (n.d.). World Values Survey Publications. Retrieved December 17, 2016, from <http://www.worldvaluessurvey.org/WVSContents.jsp>
- Ziegler, M., Poropat, A., & Mell, J. (2014). Does the length of a questionnaire matter? Expected and unexpected answers from generalizability theory. *Journal of Individual Differences*, 35(4), 250–261. <http://dx.doi.org/10.1027/1614-0001/a000147>.

Supplementary Materials:

Problems with the Big Five assessment in the World Values Survey

Contents

A: Information on variables	2
B: Data preprocessing	3
C: Figures included in main text	4
D: Descriptive statistics	8
E: Interitem correlations	9

A: Information on variables

The **o** items relate to openness, **c** items to conscientiousness, **e** items to extraversion, **a** items to agreeableness, and **s** items to emotional stability. The items with the suffix 2 are reverse coded.

Item	WVS variable	Description	Reverse coded
o1	V160J	... has an active imagination	No
o2	V160E	... has few artistic interests	Yes
c1	V160H	... does a thorough job	No
c2	V160C	... tends to be lazy	Yes
e1	V160F	... is outgoing, sociable	No
e2	V160A	... is reserved	Yes
a1	V160B	... is generally trusting	No
a2	V160G	... tends to find fault with others	Yes
s1	V160D	... is relaxed, handles stress well	No
s2	V160I	... gets nervous easily	Yes

B: Data preprocessing

```
# Load packages
library("rio")
library("ggplot2")
library("reshape2")
library("grid")
library("gridExtra")
library("tidyr")
library("stargazer")

# Load dataset (Stata format)
wvs <- import("WV6_Stata_v_2016_01_01.dta")

# Code missing values
trait.vars <- c("V160A", "V160B", "V160C", "V160D", "V160E",
               "V160F", "V160G", "V160H", "V160I", "V160J")
wvs[trait.vars][wvs[trait.vars] < 0] <- NA

# Reverse code and save variables
wvs$o1 <- wvs$V160J
wvs$o2 <- (wvs$V160E-6)*-1

wvs$c1 <- wvs$V160H
wvs$c2 <- (wvs$V160C-6)*-1

wvs$e1 <- wvs$V160F
wvs$e2 <- (wvs$V160A-6)*-1

wvs$a1 <- wvs$V160B
wvs$a2 <- (wvs$V160G-6)*-1

wvs$s1 <- wvs$V160D
wvs$s2 <- (wvs$V160I-6)*-1

wvs$male <- wvs$V240
wvs$male[wvs$male < 0] <- NA
wvs$male[wvs$male == 2] <- 0

wvs$age <- wvs$V242
wvs$age[wvs$age < 0] <- NA
```

C: Figures included in main text

```
# Make data frame with the BFI-10 items
b5 <- wvs[c("V2", "male", "age", "o1", "o2", "c1", "c2", "e1", "e2", "a1", "a2", "s1", "s2")]
b5 <- na.omit(b5)

b5.cor <- data.frame(country = unique(b5$V2),
                    n = NA,
                    cor.o = NA,
                    se.o = NA,
                    cor.c = NA,
                    se.c = NA,
                    cor.e = NA,
                    se.e = NA,
                    cor.a = NA,
                    se.a = NA,
                    cor.s = NA,
                    se.s = NA
                    )

for (i in unique(b5$V2)){
  b5.cor$n[b5.cor$country == i] <- NROW(b5[b5$V2 == i,])
  b5.cor$cor.o[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$o1,
                                                b5[b5$V2 == i,]$o2)$estimate
  b5.cor$se.o[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$o1,
                                                b5[b5$V2 == i,]$o2)$estimate /
    cor.test(b5[b5$V2 == i,]$o1, b5[b5$V2 == i,]$o2)$statistic
  b5.cor$cor.c[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$c1,
                                                b5[b5$V2 == i,]$c2)$estimate
  b5.cor$se.c[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$c1,
                                                b5[b5$V2 == i,]$c2)$estimate /
    cor.test(b5[b5$V2 == i,]$c1, b5[b5$V2 == i,]$c2)$statistic
  b5.cor$cor.e[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$e1,
                                                b5[b5$V2 == i,]$e2)$estimate
  b5.cor$se.e[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$e1,
                                                b5[b5$V2 == i,]$e2)$estimate /
    cor.test(b5[b5$V2 == i,]$e1, b5[b5$V2 == i,]$e2)$statistic
  b5.cor$cor.a[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$a1,
                                                b5[b5$V2 == i,]$a2)$estimate
  b5.cor$se.a[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$a1,
                                                b5[b5$V2 == i,]$a2)$estimate /
    cor.test(b5[b5$V2 == i,]$a1, b5[b5$V2 == i,]$a2)$statistic
  b5.cor$cor.s[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$s1,
                                                b5[b5$V2 == i,]$s2)$estimate
  b5.cor$se.s[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$s1,
                                                b5[b5$V2 == i,]$s2)$estimate /
    cor.test(b5[b5$V2 == i,]$s1, b5[b5$V2 == i,]$s2)$statistic
}
```

```

b5.cor$name <- NA
b5.cor[b5.cor$country == 12,]$name <- "Algeria"
b5.cor[b5.cor$country == 48,]$name <- "Bahrain"
b5.cor[b5.cor$country == 76,]$name <- "Brazil"
b5.cor[b5.cor$country == 156,]$name <- "China"
b5.cor[b5.cor$country == 400,]$name <- "Jordan"
b5.cor[b5.cor$country == 414,]$name <- "Kuwait"
b5.cor[b5.cor$country == 702,]$name <- "Singapore"
b5.cor[b5.cor$country == 170,]$name <- "Colombia"
b5.cor[b5.cor$country == 218,]$name <- "Ecuador"
b5.cor[b5.cor$country == 818,]$name <- "Egypt"
b5.cor[b5.cor$country == 268,]$name <- "Georgia"
b5.cor[b5.cor$country == 276,]$name <- "Germany"
b5.cor[b5.cor$country == 344,]$name <- "Hong Kong"
b5.cor[b5.cor$country == 356,]$name <- "India"
b5.cor[b5.cor$country == 368,]$name <- "Iraq"
b5.cor[b5.cor$country == 422,]$name <- "Lebanon"
b5.cor[b5.cor$country == 434,]$name <- "Libya"
b5.cor[b5.cor$country == 528,]$name <- "Netherlands"
b5.cor[b5.cor$country == 586,]$name <- "Pakistan"
b5.cor[b5.cor$country == 275,]$name <- "Palestine"
b5.cor[b5.cor$country == 646,]$name <- "Rwanda"
b5.cor[b5.cor$country == 710,]$name <- "South Africa"
b5.cor[b5.cor$country == 764,]$name <- "Thailand"
b5.cor[b5.cor$country == 788,]$name <- "Tunisia"
b5.cor[b5.cor$country == 887,]$name <- "Yemen"

fig.ii.o <- ggplot(b5.cor, aes(x = name, y=cor.o, ymin=cor.o-1.96*se.o,
                             ymax=cor.o+1.96*se.o)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Imagination, Few artistic interests)") +
  theme_minimal() +
  scale_y_continuous(breaks=c(-.5,0,.3), labels=c("-.5","0",".3")) +
  xlab("")

fig.ii.c <- ggplot(b5.cor, aes(x = name, y=cor.c, ymin=cor.c-1.96*se.c,
                             ymax=cor.c+1.96*se.c)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Not lazy, Thorough job)") +
  theme_minimal() +
  scale_y_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
  xlab("") +
  theme(axis.text.y = element_blank())

fig.ii.e <- ggplot(b5.cor, aes(x = name, y=cor.e, ymin=cor.e-1.96*se.e,
                             ymax=cor.e+1.96*se.e)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +

```



```

geom_pointrange() +
coord_flip() +
ylab("r (Not reserved, Outgoing)") +
theme_minimal() +
scale_y_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
xlab("") +
theme(axis.text.y = element_blank())

fig.ii.a <- ggplot(b5.cor, aes(x = name, y=cor.a, ymin=cor.a-1.96*se.a,
                             ymax=cor.a+1.96*se.a)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Trusting, Does not find faults)") +
  theme_minimal() +
  xlab("")

fig.ii.s <- ggplot(b5.cor, aes(x = name, y=cor.s, ymin=cor.s-1.96*se.s,
                             ymax=cor.s+1.96*se.s)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Relaxed, not nervous)") +
  scale_y_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
  theme_minimal() +
  xlab("") +
  theme(axis.text.y = element_blank())

png('figure1.png', height=8, width=8, units="in",res=700)
grid.arrange(fig.ii.o, fig.ii.c, fig.ii.e, fig.ii.a, fig.ii.s,
             widths=c(5, 4, 4), ncol=3)
dev.off()

## pdf
## 2

b5.long <- gather(b5.cor, trait, value, c(cor.o,cor.c,cor.e,cor.a,cor.s),
                 factor_key=TRUE)

png('figure2.png', height=6, width=8, units="in",res=700)
ggplot(b5.long, aes(x=value, fill=trait)) +
  geom_vline(xintercept=0, linetype="dashed") +
  geom_vline(xintercept=-0.3, colour="gray", linetype="dashed") +
  geom_vline(xintercept=0.3, colour="gray", linetype="dashed") +
  scale_y_continuous(breaks=c(0,.25,.50,.75,1), labels=c("", "", "", "", "")) +
  scale_x_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
  geom_dotplot(stackgroups = TRUE, stackratio = 1.2, binwidth=0.07,
              dotsize = 0.7, binpositions = "all") +
  scale_fill_manual("", labels = c("Openness", "Conscientiousness", "Extraversion",
                                  "Agreeableness", "Emotional Stability"),
                   values = c("#69D2E7", "#81AD99", "#C02942",
                              "#F38630", "#ECD078")) +

```

```
xlab("Item-item correlation") +  
ylab("") +  
annotate("text", x = -0.8, y = 0.27, label = "Bahrain") +  
theme_minimal()  
dev.off()
```

```
## pdf  
## 2
```

D: Descriptive statistics

```
# Get country with minimum number of observations  
min(b5.cor$n)
```

```
[1] 653
```

```
b5.cor$name[b5.cor$n == min(b5.cor$n)]
```

```
[1] "Yemen"
```

```
# Get country with maximum number of observations  
max(b5.cor$n)
```

```
[1] 3317
```

```
b5.cor$name[b5.cor$n == max(b5.cor$n)]
```

```
[1] "South Africa"
```

```
# Create summary statistics table  
stargazer(b5[c("male", "age", "o1", "o2", "c1", "c2",  
              "e1", "e2", "a1", "a2", "s1", "s2")],  
          title = "Summary statistics",  
          covariate.labels = c("Male", "Age"),  
          summary = TRUE)
```

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Tir, Feb 28, 2017 - 13:14:38

Table 2: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Male	32,880	0.494	0.500	0	1
Age	32,880	40.540	15.643	16	99
o1	32,880	3.242	1.239	1	5
o2	32,880	3.027	1.273	1	5
c1	32,880	3.617	1.248	1	5
c2	32,880	3.528	1.297	1	5
e1	32,880	3.536	1.224	1	5
e2	32,880	2.734	1.322	1	5
a1	32,880	3.374	1.295	1	5
a2	32,880	3.203	1.297	1	5
s1	32,880	3.326	1.194	1	5
s2	32,880	2.997	1.287	1	5

E: Interitem correlations

```
get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat)]<- NA
  return(cormat)
}

for(i in unique(b5$V2)) {
  cormat <- round(cor(b5[b5$V2 == i,
                      c("o1","o2","c1","c2","e1","e2","a1","a2","s1","s2")]),
                 2)
  upper_tri <- get_upper_tri(cormat)

  melted_cormat <- melt(upper_tri, na.rm = TRUE)

  p <- ggplot(data = melted_cormat, aes(Var2, Var1, fill = value))+
    geom_tile(color = "white")+
    scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                        midpoint = 0, limit = c(-1,1), space = "Lab",
                        name="Pearson\nCorrelation") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, vjust = 1,
                                      size = 12, hjust = 1))+
    coord_fixed() +
    ggtitle(b5.cor[b5.cor$country == i,]$name) +
    geom_text(aes(Var2, Var1, label = value), color = "black", size = 2) +
    theme(
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      panel.grid.major = element_blank(),
      panel.border = element_blank(),
      panel.background = element_blank(),
      axis.ticks = element_blank(),
      plot.title = element_text(size = 12),
      legend.justification = c(1, 0),
      legend.position = c(0.6, 0.7),
      legend.direction = "horizontal")+
    guides(fill = guide_colorbar(barwidth = 7, barheight = 1,
                                title.position = "top", title.hjust = 0.5))

  print(p)
}
```

