

```

# Janine Campbell
# Contact: janine.a.campbell@uis.no
#
# Date last modified: 27 May 2021
#
#####
# A) PISA, gender equality and human development over the years 2006 - 2018
#####
#
# This study explores:
#
# 1) The relationship between Gender Inequality Index scores and Human
# Development Index scores (both from UNDP) with PISA scores for each
# PISA testing cycle, 2006-2018.
# DATASET: PISA_complete_data
#
# 2) The relationship between change in gender equality and human
# development, and change in PISA scores, 2006-2018.
# DATASET: PISA_DATA_TIME
#
# 3) The relationship between the items of gender equality and human
# development 2018, and PISA scores 2018.
# DATASET: PISA2018_HDI_GII
#
# 4) The relationship between change in the items of gender equality and
# human development, and change in PISA scores, 2006-2018.
# DATASET: PISA_HDI_GII_CHANGE
#
#-----
# Remove all objects currently stored in active memory
#-----
rm(list=ls(all=TRUE))
# rm means remove, and it can also be used for an individual condition/variable
#-----
# Check / set working directory
#-----
getwd()
# Check that working directory is correct. To change use following code:
# setwd("/Users/xxxxxxx")
options(prompt="R> ")
# This is just a cosmetic change to make the prompt more apparent
#-----
# Install and load packages for session
#-----
#
# Package names
packages <- c("ggplot2", "readxl", "dplyr", "tidyr", "ggpubr", "Hmisc", "plyr",
  "tidyverse", "jtools", "psych", "sjstats", "rcompanion")

# Install packages not yet installed
installed_packages <- packages %in% rownames(installed.packages())
if (any(installed_packages == FALSE)) {
  install.packages(packages[!installed_packages])
}

# Packages loading
invisible(lapply(packages, library, character.only = TRUE))

# function for making simple regression plots
ggplotRegression <- function(fit){ggplot(fit$model,
  aes_string(x = names(fit$model)[2],
    y = names(fit$model)[1])) +

  geom_point() +
  stat_smooth(method = "lm", col = "red") +
  labs(title = paste("R2 = ",signif(summary(fit)$r.squared, 5),
    "Int. =",signif(fit$coef[[1]], 5),
    "Slope =",signif(fit$coef[[2]], 5),
    "P =",signif(summary(fit)$coef[2,4], 5)))

#
# Remove scientific notation
options(scipen=999)
# Limit results to 6 decimal places
options(digits=6)
#
#-----
# 1) The relationship between Gender Inequality Index scores and Human

```

```

# Development Index scores (both from UNDP) with PISA scores for each
# PISA testing cycle, 2006-2018.
#-----
#
# Load data set
#
DATA <- read_excel("PISA_complete_data.xlsx")
#
head(DATA)
#
#####
# Note on cases:
# A case in these analyses is an instance of country + test. For example,
# <<Argentina 2006 Math>> is one case, and <<Argentina 2006 Reading>> is another
# case. An additional case <<Argentina 2006 Average>> is the average score for the
# country. All cases have data for the corresponding year from the Gender
# Inequality Index (inverted to represent gender equality GENEQ) and the
# Human Development Index (HDI).
#
#####
# Description of variables:
# PISA: PISA score (for that country, year, test combination)
# GENEQ: Gender Equality Index score (inverse of GII)
# DEVEL: Human Development Index score
# HIGHPISA: factor 1 = PISA score 500 or above
# LOWPISA: factor 1 = PISA score 420 or below
# HIGHGENEQ: factor 1 = better than the OECD average GII score for that year
# HIGHDEV: factor 1 = better than the OECD average HDI score for that year
#
#####
# Make separate dataframes for average, reading, math, science
#####
#
DATA_AV <- DATA[which(DATA$SUBJECT=="Average"),]
DATA_RE <- DATA[which(DATA$SUBJECT=="Reading"),]
DATA_MA <- DATA[which(DATA$SUBJECT=="Math"),]
DATA_SC <- DATA[which(DATA$SUBJECT=="Science"),]
#
#####
# 1. Analysis for AVERAGE
#####
#
#-----
# Summary data (n=286)
#-----
#
# Descriptive statistics PISA scores
summary(DATA_AV$PISA)
sd(DATA_AV$PISA)
hist(DATA_AV$PISA)
#
# Descriptive statistics Gender Equality
summary(DATA_AV$GENEQ)
sd(DATA_AV$GENEQ)
hist(DATA_AV$GENEQ)
#
# Descriptive statistics Human Development
summary(DATA_AV$DEVEL)
sd(DATA_AV$DEVEL)
hist(DATA_AV$DEVEL)
#
#-----
# Subset data for correlation and quick regression - all data
#-----
#
ALLCORR_AV <- DATA_AV[,c("PISA", "GENEQ", "DEVEL")]
#
rcorr(as.matrix(ALLCORR_AV))
#
lmALLG_AV <- lm(ALLCORR_AV$PISA ~ ALLCORR_AV$GENEQ)
summ(lmALLG_AV)
ggplotRegression(lmALLG_AV)
#
lmALLD_AV <- lm(ALLCORR_AV$PISA ~ ALLCORR_AV$DEVEL)
summ(lmALLD_AV)
ggplotRegression(lmALLD_AV)

```

```

#
#-----
# Creat contingency tables and test for independence
#-----
#
# Contingency table High PISA, HIGENEQ
prop.table(table(DATA_AV$HIGENEQ, DATA_AV$HIGHPIISA), margin=2)*100
chisq.test(DATA_AV$HIGENEQ, DATA_AV$HIGHPIISA)
chisq.test(DATA_AV$HIGENEQ, DATA_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIGENEQ
prop.table(table(DATA_AV$HIGENEQ, DATA_AV$LOWPIISA), margin=2)*100
chisq.test(DATA_AV$HIGENEQ, DATA_AV$LOWPIISA)
chisq.test(DATA_AV$HIGENEQ, DATA_AV$LOWPIISA, correct = F)
#
# Contingency table High PISA, HIDEV
prop.table(table(DATA_AV$HIDEV, DATA_AV$HIGHPIISA), margin=2)*100
chisq.test(DATA_AV$HIDEV, DATA_AV$HIGHPIISA)
chisq.test(DATA_AV$HIDEV, DATA_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIDEV
prop.table(table(DATA_AV$HIDEV, DATA_AV$LOWPIISA), margin=2)*100
chisq.test(DATA_AV$HIDEV, DATA_AV$LOWPIISA)
chisq.test(DATA_AV$HIDEV, DATA_AV$LOWPIISA, correct = F)
#
#-----
# Subset data for correlations - 2006 - n=48
#-----
#
CORR2006_AV <- DATA_AV[which(DATA_AV$YEAR=="2006"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2006_AV$PISA)
sd(CORR2006_AV$PISA)
#
summary(CORR2006_AV$GENEQ)
sd(CORR2006_AV$GENEQ)
#
summary(CORR2006_AV$DEVEL)
sd(CORR2006_AV$DEVEL)
#
rcorr(as.matrix(CORR2006_AV))
#
# Quick visualization of the relationships
lm2006G_AV <- lm(CORR2006_AV$PISA ~ CORR2006_AV$GENEQ)
summ(lm2006G_AV)
ggplotRegression(lm2006G_AV)
#
lm2006D_AV <- lm(CORR2006_AV$PISA ~ CORR2006_AV$DEVEL)
summ(lm2006D_AV)
ggplotRegression(lm2006D_AV)
#
#-----
# Contingency tables and test for independence
#-----
#
CONT2006_AV <- DATA_AV[which(DATA_AV$YEAR=="2006"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2006_AV$HIGENEQ, CONT2006_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2006_AV$HIGENEQ, CONT2006_AV$HIGHPIISA)
chisq.test(CONT2006_AV$HIGENEQ, CONT2006_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2006_AV$HIGENEQ, CONT2006_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2006_AV$HIGENEQ, CONT2006_AV$LOWPIISA)
chisq.test(CONT2006_AV$HIGENEQ, CONT2006_AV$LOWPIISA, correct = F)
#
# Contingency table High PISA, HIDEV
prop.table(table(CONT2006_AV$HIDEV, CONT2006_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2006_AV$HIDEV, CONT2006_AV$HIGHPIISA)
chisq.test(CONT2006_AV$HIDEV, CONT2006_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2006_AV$HIDEV, CONT2006_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2006_AV$HIDEV, CONT2006_AV$LOWPIISA)
chisq.test(CONT2006_AV$HIDEV, CONT2006_AV$LOWPIISA, correct = F)
#
#-----
# Subset data for correlations - 2009 - N=58

```

```

#-----
#
CORR2009_AV <- DATA_AV[which(DATA_AV$YEAR=="2009"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2009_AV$PISA)
sd(CORR2009_AV$PISA)
#
summary(CORR2009_AV$GENEQ)
sd(CORR2009_AV$GENEQ)
#
summary(CORR2009_AV$DEVEL)
sd(CORR2009_AV$DEVEL)
#
rcorr(as.matrix(CORR2009_AV))
#
# Quick visualization of relationships
lm2009G_AV <- lm(CORR2009_AV$PISA ~ CORR2009_AV$GENEQ)
summ(lm2009G_AV)
ggplotRegression(lm2009G_AV)
#
lm2009D_AV <- lm(CORR2009_AV$PISA ~ CORR2009_AV$DEVEL)
summ(lm2009D_AV)
ggplotRegression(lm2009D_AV)
#
#-----
# Contingency tables and test for independence
#-----
#
CONT2009_AV <- DATA_AV[which(DATA_AV$YEAR=="2009"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2009_AV$HIGENEQ, CONT2009_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2009_AV$HIGENEQ, CONT2009_AV$HIGHPIISA)
chisq.test(CONT2009_AV$HIGENEQ, CONT2009_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2009_AV$HIGENEQ, CONT2009_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2009_AV$HIGENEQ, CONT2009_AV$LOWPIISA)
chisq.test(CONT2009_AV$HIGENEQ, CONT2009_AV$LOWPIISA, correct = F)
#
# Contingency table High PISA, HIDEV
prop.table(table(CONT2009_AV$HIDEV, CONT2009_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2009_AV$HIDEV, CONT2009_AV$HIGHPIISA)
chisq.test(CONT2009_AV$HIDEV, CONT2009_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2009_AV$HIDEV, CONT2009_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2009_AV$HIDEV, CONT2009_AV$LOWPIISA)
chisq.test(CONT2009_AV$HIDEV, CONT2009_AV$LOWPIISA, correct = F)
#
#-----
# Subset data for correlations - 2012 - N=57
#-----
#
CORR2012_AV <- DATA_AV[which(DATA_AV$YEAR=="2012"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2012_AV$PISA)
sd(CORR2012_AV$PISA)
#
summary(CORR2012_AV$GENEQ)
sd(CORR2012_AV$GENEQ)
#
summary(CORR2012_AV$DEVEL)
sd(CORR2012_AV$DEVEL)
#
rcorr(as.matrix(CORR2012_AV))
#
# Quick visualization of relationships
lm2012G_AV <- lm(CORR2012_AV$PISA ~ CORR2012_AV$GENEQ)
summ(lm2012G_AV)
ggplotRegression(lm2012G_AV)
#
lm2012D_AV <- lm(CORR2012_AV$PISA ~ CORR2012_AV$DEVEL)
summ(lm2012D_AV)
ggplotRegression(lm2012D_AV)
#
#-----

```

```

# Contingency tables and test for independence
#-----
#
CONT2012_AV <- DATA_AV[which(DATA_AV$YEAR=="2012"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2012_AV$HIGENEQ, CONT2012_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2012_AV$HIGENEQ, CONT2012_AV$HIGHPIISA)
chisq.test(CONT2012_AV$HIGENEQ, CONT2012_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2012_AV$HIGENEQ, CONT2012_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2012_AV$HIGENEQ, CONT2012_AV$LOWPIISA)
chisq.test(CONT2012_AV$HIGENEQ, CONT2012_AV$LOWPIISA, correct = F)
#
# Contingency table High PISA, HIDEV
prop.table(table(CONT2012_AV$HIDEV, CONT2012_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2012_AV$HIDEV, CONT2012_AV$HIGHPIISA)
chisq.test(CONT2012_AV$HIDEV, CONT2012_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2012_AV$HIDEV, CONT2012_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2012_AV$HIDEV, CONT2012_AV$LOWPIISA)
chisq.test(CONT2012_AV$HIDEV, CONT2012_AV$LOWPIISA, correct = F)
#
#-----
# Subset data for correlations - 2015 - N=60
#-----
#
CORR2015_AV <- DATA_AV[which(DATA_AV$YEAR=="2015"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2015_AV$PISA)
sd(CORR2015_AV$PISA)
#
summary(CORR2015_AV$GENEQ)
sd(CORR2015_AV$GENEQ)
#
summary(CORR2015_AV$DEVEL)
sd(CORR2015_AV$DEVEL)
#
rcorr(as.matrix(CORR2015_AV))
#
# Quick visualization of relationships
lm2015G_AV <- lm(CORR2015_AV$PISA ~ CORR2015_AV$GENEQ)
summ(lm2015G_AV)
ggplotRegression(lm2015G_AV)
#
lm2015D_AV <- lm(CORR2015_AV$PISA ~ CORR2015_AV$DEVEL)
summ(lm2015D_AV)
ggplotRegression(lm2015D_AV)
#
#-----
# Contingency tables and test for independence
#-----
#
CONT2015_AV <- DATA_AV[which(DATA_AV$YEAR=="2015"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2015_AV$HIGENEQ, CONT2015_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2015_AV$HIGENEQ, CONT2015_AV$HIGHPIISA)
chisq.test(CONT2015_AV$HIGENEQ, CONT2015_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2015_AV$HIGENEQ, CONT2015_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2015_AV$HIGENEQ, CONT2015_AV$LOWPIISA)
chisq.test(CONT2015_AV$HIGENEQ, CONT2015_AV$LOWPIISA, correct = F)
#
# Contingency table High PISA, HIDEV
prop.table(table(CONT2015_AV$HIDEV, CONT2015_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2015_AV$HIDEV, CONT2015_AV$HIGHPIISA)
chisq.test(CONT2015_AV$HIDEV, CONT2015_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2015_AV$HIDEV, CONT2015_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2015_AV$HIDEV, CONT2015_AV$LOWPIISA)
chisq.test(CONT2015_AV$HIDEV, CONT2015_AV$LOWPIISA, correct = F)
#
#-----

```

```

# Subset data for correlations - 2018 - N=62
#-----
#
CORR2018_AV <- DATA_AV[which(DATA_AV$YEAR=="2018"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2018_AV$PISA)
sd(CORR2018_AV$PISA)
#
summary(CORR2018_AV$GENEQ)
sd(CORR2018_AV$GENEQ)
#
summary(CORR2018_AV$DEVEL)
sd(CORR2018_AV$DEVEL)
#
rcorr(as.matrix(CORR2018_AV))
#
# Quick visualization of relationships
lm2018G_AV <- lm(CORR2018_AV$PISA ~ CORR2018_AV$GENEQ)
summ(lm2018G_AV)
ggplotRegression(lm2018G_AV)
#
lm2018D_AV <- lm(CORR2018_AV$PISA ~ CORR2018_AV$DEVEL)
summ(lm2018D_AV)
ggplotRegression(lm2018D_AV)
#
#-----
# Contingency tables and test for independence
#-----
#
CONT2018_AV <- DATA_AV[which(DATA_AV$YEAR=="2018"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2018_AV$HIGENEQ, CONT2018_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2018_AV$HIGENEQ, CONT2018_AV$HIGHPIISA)
chisq.test(CONT2018_AV$HIGENEQ, CONT2018_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2018_AV$HIGENEQ, CONT2018_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2018_AV$HIGENEQ, CONT2018_AV$LOWPIISA)
chisq.test(CONT2018_AV$HIGENEQ, CONT2018_AV$LOWPIISA, correct = F)
#
# Contingency table High PISA, HIDEV
prop.table(table(CONT2018_AV$HIDEV, CONT2018_AV$HIGHPIISA), margin=2)*100
chisq.test(CONT2018_AV$HIDEV, CONT2018_AV$HIGHPIISA)
chisq.test(CONT2018_AV$HIDEV, CONT2018_AV$HIGHPIISA, correct = F)
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2018_AV$HIDEV, CONT2018_AV$LOWPIISA), margin=2)*100
chisq.test(CONT2018_AV$HIDEV, CONT2018_AV$LOWPIISA)
chisq.test(CONT2018_AV$HIDEV, CONT2018_AV$LOWPIISA, correct = F)
#
#####
# 2. Analysis for READING
#####
#
#-----
# Summary data (n=274)
#-----
#
# Descriptive statistics PISA scores
summary(DATA_RE$PISA)
sd(DATA_RE$PISA)
hist(DATA_RE$PISA)
#
# Descriptive statistics Gender Equality
summary(DATA_RE$GENEQ)
sd(DATA_RE$GENEQ)
hist(DATA_RE$GENEQ)
#
# Descriptive statistics Human Development
summary(DATA_RE$DEVEL)
sd(DATA_RE$DEVEL)
hist(DATA_RE$DEVEL)
#
#-----
# Subset data for correlation and quick regression - all data
#-----

```

```

#
ALLCORR_RE <- DATA_RE[,c("PISA", "GENEQ", "DEVEL")]
#
rcorr(as.matrix(ALLCORR_RE))
#
lmALLG_RE <- lm(ALLCORR_RE$PISA ~ ALLCORR_RE$GENEQ)
summ(lmALLG_RE)
ggplotRegression(lmALLG_RE)
#
lmALLD_RE <- lm(ALLCORR_RE$PISA ~ ALLCORR_RE$DEVEL)
summ(lmALLD_RE)
ggplotRegression(lmALLD_RE)
#
#-----
# Contingency tables
#-----
#
# Contingency table High PISA, HIGENEQ
prop.table(table(DATA_RE$HIGENEQ, DATA_RE$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(DATA_RE$HIGENEQ, DATA_RE$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(DATA_RE$HIDEV, DATA_RE$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(DATA_RE$HIDEV, DATA_RE$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2006 - n=46
#-----
#
CORR2006_RE <- DATA_RE[which(DATA_RE$YEAR=="2006"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2006_RE$PISA)
sd(CORR2006_RE$PISA)
#
summary(CORR2006_RE$GENEQ)
sd(CORR2006_RE$GENEQ)
#
summary(CORR2006_RE$DEVEL)
sd(CORR2006_RE$DEVEL)
#
rcorr(as.matrix(CORR2006_RE))
#
# Quick visualization of the relationships
lm2006G_RE <- lm(CORR2006_RE$PISA ~ CORR2006_RE$GENEQ)
summ(lm2006G_RE)
ggplotRegression(lm2006G_RE)
#
lm2006D_RE <- lm(CORR2006_RE$PISA ~ CORR2006_RE$DEVEL)
summ(lm2006D_RE)
ggplotRegression(lm2006D_RE)
#
#-----
# Contingency tables
#-----
#
CONT2006_RE <- DATA_RE[which(DATA_RE$YEAR=="2006"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2006_RE$HIGENEQ, CONT2006_RE$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2006_RE$HIGENEQ, CONT2006_RE$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2006_RE$HIDEV, CONT2006_RE$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2006_RE$HIDEV, CONT2006_RE$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2009 - N=55
#-----
#
CORR2009_RE <- DATA_RE[which(DATA_RE$YEAR=="2009"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2009_RE$PISA)
sd(CORR2009_RE$PISA)

```

```

#
summary(CORR2009_RE$GENEQ)
sd(CORR2009_RE$GENEQ)
#
summary(CORR2009_RE$DEVEL)
sd(CORR2009_RE$DEVEL)
#
rcorr(as.matrix(CORR2009_RE))
#
# Quick visualization of relationships
lm2009G_RE <- lm(CORR2009_RE$PISA ~ CORR2009_RE$GENEQ)
summ(lm2009G_RE)
ggplotRegression(lm2009G_RE)
#
lm2009D_RE <- lm(CORR2009_RE$PISA ~ CORR2009_RE$DEVEL)
summ(lm2009D_RE)
ggplotRegression(lm2009D_RE)
#
#-----
# Contingency tables
#-----
#
CONT2009_RE <- DATA_RE[which(DATA_RE$YEAR=="2009"),c("HIGH PISA", "LOW PISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2009_RE$HIGENEQ, CONT2009_RE$HIGH PISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2009_RE$HIGENEQ, CONT2009_RE$LOW PISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2009_RE$HIDEV, CONT2009_RE$HIGH PISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2009_RE$HIDEV, CONT2009_RE$LOW PISA), margin=2)*100
#
#-----
# Subset data for correlations - 2012 - N=55
#-----
#
CORR2012_RE <- DATA_RE[which(DATA_RE$YEAR=="2012"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2012_RE$PISA)
sd(CORR2012_RE$PISA)
#
summary(CORR2012_RE$GENEQ)
sd(CORR2012_RE$GENEQ)
#
summary(CORR2012_RE$DEVEL)
sd(CORR2012_RE$DEVEL)
#
rcorr(as.matrix(CORR2012_RE))
#
# Quick visualization of relationships
lm2012G_RE <- lm(CORR2012_RE$PISA ~ CORR2012_RE$GENEQ)
summ(lm2012G_RE)
ggplotRegression(lm2012G_RE)
#
lm2012D_RE <- lm(CORR2012_RE$PISA ~ CORR2012_RE$DEVEL)
summ(lm2012D_RE)
ggplotRegression(lm2012D_RE)
#
#-----
# Contingency tables
#-----
#
CONT2012_RE <- DATA_RE[which(DATA_RE$YEAR=="2012"),c("HIGH PISA", "LOW PISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2012_RE$HIGENEQ, CONT2012_RE$HIGH PISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2012_RE$HIGENEQ, CONT2012_RE$LOW PISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2012_RE$HIDEV, CONT2012_RE$HIGH PISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2012_RE$HIDEV, CONT2012_RE$LOW PISA), margin=2)*100
#

```



```

#-----
# Subset data for correlations - 2015 - N=58
#-----
#
CORR2015_RE <- DATA_RE[which(DATA_RE$YEAR=="2015"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2015_RE$PISA)
sd(CORR2015_RE$PISA)
#
summary(CORR2015_RE$GENEQ)
sd(CORR2015_RE$GENEQ)
#
summary(CORR2015_RE$DEVEL)
sd(CORR2015_RE$DEVEL)
#
rcorr(as.matrix(CORR2015_RE))
#
# Quick visualization of relationships
lm2015G_RE <- lm(CORR2015_RE$PISA ~ CORR2015_RE$GENEQ)
summ(lm2015G_RE)
ggplotRegression(lm2015G_RE)
#
lm2015D_RE <- lm(CORR2015_RE$PISA ~ CORR2015_RE$DEVEL)
summ(lm2015D_RE)
ggplotRegression(lm2015D_RE)
#
#-----
# Contingency tables
#-----
#
CONT2015_RE <- DATA_RE[which(DATA_RE$YEAR=="2015"),c("HIGH PISA", "LOW PISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2015_RE$HIGENEQ, CONT2015_RE$HIGH PISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2015_RE$HIGENEQ, CONT2015_RE$LOW PISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2015_RE$HIDEV, CONT2015_RE$HIGH PISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2015_RE$HIDEV, CONT2015_RE$LOW PISA), margin=2)*100
#
#-----
# Subset data for correlations - 2018 - N=60
#-----
#
CORR2018_RE <- DATA_RE[which(DATA_RE$YEAR=="2018"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2018_RE$PISA)
sd(CORR2018_RE$PISA)
#
summary(CORR2018_RE$GENEQ)
sd(CORR2018_RE$GENEQ)
#
summary(CORR2018_RE$DEVEL)
sd(CORR2018_RE$DEVEL)
#
rcorr(as.matrix(CORR2018_RE))
#
# Quick visualization of relationships
lm2018G_RE <- lm(CORR2018_RE$PISA ~ CORR2018_RE$GENEQ)
summ(lm2018G_RE)
ggplotRegression(lm2018G_RE)
#
lm2018D_RE <- lm(CORR2018_RE$PISA ~ CORR2018_RE$DEVEL)
summ(lm2018D_RE)
ggplotRegression(lm2018D_RE)
#
#-----
# Contingency tables
#-----
#
CONT2018_RE <- DATA_RE[which(DATA_RE$YEAR=="2018"),c("HIGH PISA", "LOW PISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ

```

```

prop.table(table(CONT2018_RE$HIGENEQ, CONT2018_RE$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2018_RE$HIGENEQ, CONT2018_RE$LOWPIA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2018_RE$HIDEV, CONT2018_RE$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2018_RE$HIDEV, CONT2018_RE$LOWPIA), margin=2)*100
#
#####
# 3. Analysis for MATH
#####
#
#-----
# Summary data (n=276)
#-----
#
# Descriptive statistics PISA scores
summary(DATA_MAS$PISA)
sd(DATA_MAS$PISA)
hist(DATA_MAS$PISA)
#
# Descriptive statistics Gender Equality
summary(DATA_MAS$GENEQ)
sd(DATA_MAS$GENEQ)
hist(DATA_MAS$GENEQ)
#
# Descriptive statistics Humand Development
summary(DATA_MAS$DEVEL)
sd(DATA_MAS$DEVEL)
hist(DATA_MAS$DEVEL)
#
#-----
# Subset data for correlation and quick regression - all data
#-----
#
ALLCORR_MA <- DATA_MA[,c("PISA", "GENEQ", "DEVEL")]
#
rcorr(as.matrix(ALLCORR_MA))
#
lmALLG_MA <- lm(ALLCORR_MAS$PISA ~ ALLCORR_MAS$GENEQ)
summ(lmALLG_MA)
ggplotRegression(lmALLG_MA)
#
lmALLD_MA <- lm(ALLCORR_MAS$PISA ~ ALLCORR_MAS$DEVEL)
summ(lmALLD_MA)
ggplotRegression(lmALLD_MA)
#
#-----
# Contingency tables
#-----
#
# Contingency table High PISA, HIGENEQ
prop.table(table(DATA_MAS$HIGENEQ, DATA_MAS$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(DATA_MAS$HIGENEQ, DATA_MAS$LOWPIA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(DATA_MAS$HIDEV, DATA_MAS$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(DATA_MAS$HIDEV, DATA_MAS$LOWPIA), margin=2)*100
#
#-----
# Subset data for correlations - 2006 - n=47
#-----
#
CORR2006_MA <- DATA_MA[which(DATA_MAS$YEAR=="2006"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2006_MAS$PISA)
sd(CORR2006_MAS$PISA)
#
summary(CORR2006_MAS$GENEQ)
sd(CORR2006_MAS$GENEQ)
#
summary(CORR2006_MAS$DEVEL)
sd(CORR2006_MAS$DEVEL)
#
rcorr(as.matrix(CORR2006_MA))

```

```

#
# Quick visualization of the relationships
lm2006G_MA <- lm(CORR2006_MA$PISA ~ CORR2006_MA$GENEQ)
summ(lm2006G_MA)
ggplotRegression(lm2006G_MA)
#
lm2006D_MA <- lm(CORR2006_MA$PISA ~ CORR2006_MA$DEVEL)
summ(lm2006D_MA)
ggplotRegression(lm2006D_MA)
#
#-----
# Contingency tables
#-----
#
CONT2006_MA <- DATA_MA[which(DATA_MA$YEAR=="2006"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2006_MA$HIGENEQ, CONT2006_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2006_MA$HIGENEQ, CONT2006_MA$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2006_MA$HIDEV, CONT2006_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2006_MA$HIDEV, CONT2006_MA$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2009 - N=55
#-----
#
CORR2009_MA <- DATA_MA[which(DATA_MA$YEAR=="2009"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2009_MA$PISA)
sd(CORR2009_MA$PISA)
#
summary(CORR2009_MA$GENEQ)
sd(CORR2009_MA$GENEQ)
#
summary(CORR2009_MA$DEVEL)
sd(CORR2009_MA$DEVEL)
#
rcorr(as.matrix(CORR2009_MA))
#
# Quick visualization of relationships
lm2009G_MA <- lm(CORR2009_MA$PISA ~ CORR2009_MA$GENEQ)
summ(lm2009G_MA)
ggplotRegression(lm2009G_MA)
#
lm2009D_MA <- lm(CORR2009_MA$PISA ~ CORR2009_MA$DEVEL)
summ(lm2009D_MA)
ggplotRegression(lm2009D_MA)
#
#-----
# Contingency tables
#-----
#
CONT2009_MA <- DATA_MA[which(DATA_MA$YEAR=="2009"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2009_MA$HIGENEQ, CONT2009_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2009_MA$HIGENEQ, CONT2009_MA$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2009_MA$HIDEV, CONT2009_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2009_MA$HIDEV, CONT2009_MA$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2012 - N=55
#-----
#
CORR2012_MA <- DATA_MA[which(DATA_MA$YEAR=="2012"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2012_MA$PISA)
sd(CORR2012_MA$PISA)

```

```

#
summary(CORR2012_MA$GENEQ)
sd(CORR2012_MA$GENEQ)
#
summary(CORR2012_MA$DEVEL)
sd(CORR2012_MA$DEVEL)
#
rcorr(as.matrix(CORR2012_MA))
#
# Quick visualization of relationships
lm2012G_MA <- lm(CORR2012_MA$PISA ~ CORR2012_MA$GENEQ)
summ(lm2012G_MA)
ggplotRegression(lm2012G_MA)
#
lm2012D_MA <- lm(CORR2012_MA$PISA ~ CORR2012_MA$DEVEL)
summ(lm2012D_MA)
ggplotRegression(lm2012D_MA)
#
#-----
# Contingency tables
#-----
#
CONT2012_MA <- DATA_MA[which(DATA_MA$YEAR=="2012"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2012_MA$HIGENEQ, CONT2012_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2012_MA$HIGENEQ, CONT2012_MA$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2012_MA$HIDEV, CONT2012_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2012_MA$HIDEV, CONT2012_MA$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2015 - N=60
#-----
#
CORR2015_MA <- DATA_MA[which(DATA_MA$YEAR=="2015"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2015_MA$PISA)
sd(CORR2015_MA$PISA)
#
summary(CORR2015_MA$GENEQ)
sd(CORR2015_MA$GENEQ)
#
summary(CORR2015_MA$DEVEL)
sd(CORR2015_MA$DEVEL)
#
rcorr(as.matrix(CORR2015_MA))
#
# Quick visualization of relationships
lm2015G_MA <- lm(CORR2015_MA$PISA ~ CORR2015_MA$GENEQ)
summ(lm2015G_MA)
ggplotRegression(lm2015G_MA)
#
lm2015D_MA <- lm(CORR2015_MA$PISA ~ CORR2015_MA$DEVEL)
summ(lm2015D_MA)
ggplotRegression(lm2015D_MA)
#
#-----
# Contingency tables
#-----
#
CONT2015_MA <- DATA_MA[which(DATA_MA$YEAR=="2015"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2015_MA$HIGENEQ, CONT2015_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2015_MA$HIGENEQ, CONT2015_MA$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2015_MA$HIDEV, CONT2015_MA$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2015_MA$HIDEV, CONT2015_MA$LOWPIISA), margin=2)*100
#

```

```

#-----
# Subset data for correlations - 2018 - N=59
#-----
#
CORR2018_MA <- DATA_MA[which(DATA_MA$YEAR=="2018"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2018_MA$PISA)
sd(CORR2018_MA$PISA)
#
summary(CORR2018_MA$GENEQ)
sd(CORR2018_MA$GENEQ)
#
summary(CORR2018_MA$DEVEL)
sd(CORR2018_MA$DEVEL)
#
rcorr(as.matrix(CORR2018_MA))
#
# Quick visualization of relationships
lm2018G_MA <- lm(CORR2018_MA$PISA ~ CORR2018_MA$GENEQ)
summ(lm2018G_MA)
ggplotRegression(lm2018G_MA)
#
lm2018D_MA <- lm(CORR2018_MA$PISA ~ CORR2018_MA$DEVEL)
summ(lm2018D_MA)
ggplotRegression(lm2018D_MA)
#
#-----
# Contingency tables
#-----
#
CONT2018_MA <- DATA_MA[which(DATA_MA$YEAR=="2018"),c("HIGHPIA", "LOWPIA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2018_MA$HIGENEQ, CONT2018_MA$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2018_MA$HIGENEQ, CONT2018_MA$LOWPIA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2018_MA$HIDEV, CONT2018_MA$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2018_MA$HIDEV, CONT2018_MA$LOWPIA), margin=2)*100
#
#####
# 4. Analysis for SCIENCE
#####
#
#-----
# Summary data (n=281)
#-----
#
# Descriptive statistics PISA scores
summary(DATA_SC$PISA)
sd(DATA_SC$PISA)
hist(DATA_SC$PISA)
#
# Descriptive statistics Gender Equality
summary(DATA_SC$GENEQ)
sd(DATA_SC$GENEQ)
hist(DATA_SC$GENEQ)
#
# Descriptive statistics Human Development
summary(DATA_SC$DEVEL)
sd(DATA_SC$DEVEL)
hist(DATA_SC$DEVEL)
#
#-----
# Subset data for correlation and quick regression - all data
#-----
#
ALLCORR_SC <- DATA_SC[,c("PISA", "GENEQ", "DEVEL")]
#
rcorr(as.matrix(ALLCORR_SC))
#
lmALLG_SC <- lm(ALLCORR_SC$PISA ~ ALLCORR_SC$GENEQ)
summ(lmALLG_SC)
ggplotRegression(lmALLG_SC)

```

```

#
lmALLD_SC <- lm(ALLCORR_SC$PISA ~ ALLCORR_SC$DEVEL)
summ(lmALLD_SC)
ggplotRegression(lmALLD_SC)
#
#-----
# Contingency tables
#-----
#
# Contingency table High PISA, HIGENEQ
prop.table(table(DATA_SC$HIGENEQ, DATA_SC$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(DATA_SC$HIGENEQ, DATA_SC$LOWPIA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(DATA_SC$HIDEV, DATA_SC$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(DATA_SC$HIDEV, DATA_SC$LOWPIA), margin=2)*100
#
#-----
# Subset data for correlations - 2006 - n=48
#-----
#
CORR2006_SC <- DATA_SC[which(DATA_SC$YEAR=="2006"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2006_SC$PISA)
sd(CORR2006_SC$PISA)
#
summary(CORR2006_SC$GENEQ)
sd(CORR2006_SC$GENEQ)
#
summary(CORR2006_SC$DEVEL)
sd(CORR2006_SC$DEVEL)
#
rcorr(as.matrix(CORR2006_SC))
#
# Quick visualization of the relationships
lm2006G_SC <- lm(CORR2006_SC$PISA ~ CORR2006_SC$GENEQ)
summ(lm2006G_SC)
ggplotRegression(lm2006G_SC)
#
lm2006D_SC <- lm(CORR2006_SC$PISA ~ CORR2006_SC$DEVEL)
summ(lm2006D_SC)
ggplotRegression(lm2006D_SC)
#
#-----
# Contingency tables
#-----
#
CONT2006_SC <- DATA_SC[which(DATA_SC$YEAR=="2006"),c("HIGHPIA", "LOWPIA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2006_SC$HIGENEQ, CONT2006_SC$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2006_SC$HIGENEQ, CONT2006_SC$LOWPIA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2006_SC$HIDEV, CONT2006_SC$HIGHPIA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2006_SC$HIDEV, CONT2006_SC$LOWPIA), margin=2)*100
#
#-----
# Subset data for correlations - 2009 - N=56
#-----
#
CORR2009_SC <- DATA_SC[which(DATA_SC$YEAR=="2009"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2009_SC$PISA)
sd(CORR2009_SC$PISA)
#
summary(CORR2009_SC$GENEQ)
sd(CORR2009_SC$GENEQ)
#
summary(CORR2009_SC$DEVEL)
sd(CORR2009_SC$DEVEL)
#
rcorr(as.matrix(CORR2009_SC))

```

```

#
# Quick visualization of relationships
lm2009G_SC <- lm(CORR2009_SC$PISA ~ CORR2009_SC$GENEQ)
summ(lm2009G_SC)
ggplotRegression(lm2009G_SC)
#
lm2009D_SC <- lm(CORR2009_SC$PISA ~ CORR2009_SC$DEVEL)
summ(lm2009D_SC)
ggplotRegression(lm2009D_SC)
#
#-----
# Contingency tables
#-----
#
CONT2009_SC <- DATA_SC[which(DATA_SC$YEAR=="2009"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2009_SC$HIGENEQ, CONT2009_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2009_SC$HIGENEQ, CONT2009_SC$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2009_SC$HIDEV, CONT2009_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2009_SC$HIDEV, CONT2009_SC$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2012 - N=56
#-----
#
CORR2012_SC <- DATA_SC[which(DATA_SC$YEAR=="2012"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2012_SC$PISA)
sd(CORR2012_SC$PISA)
#
summary(CORR2012_SC$GENEQ)
sd(CORR2012_SC$GENEQ)
#
summary(CORR2012_SC$DEVEL)
sd(CORR2012_SC$DEVEL)
#
rcorr(as.matrix(CORR2012_SC))
#
# Quick visualization of relationships
lm2012G_SC <- lm(CORR2012_SC$PISA ~ CORR2012_SC$GENEQ)
summ(lm2012G_SC)
ggplotRegression(lm2012G_SC)
#
lm2012D_SC <- lm(CORR2012_SC$PISA ~ CORR2012_SC$DEVEL)
summ(lm2012D_SC)
ggplotRegression(lm2012D_SC)
#
#-----
# Contingency tables
#-----
#
CONT2012_SC <- DATA_SC[which(DATA_SC$YEAR=="2012"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2012_SC$HIGENEQ, CONT2012_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2012_SC$HIGENEQ, CONT2012_SC$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2012_SC$HIDEV, CONT2012_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2012_SC$HIDEV, CONT2012_SC$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2015 - N=59
#-----
#
CORR2015_SC <- DATA_SC[which(DATA_SC$YEAR=="2015"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2015_SC$PISA)
sd(CORR2015_SC$PISA)

```

```

#
summary(CORR2015_SC$GENEQ)
sd(CORR2015_SC$GENEQ)
#
summary(CORR2015_SC$DEVEL)
sd(CORR2015_SC$DEVEL)
#
rcorr(as.matrix(CORR2015_SC))
#
# Quick visualization of relationships
lm2015G_SC <- lm(CORR2015_SC$PISA ~ CORR2015_SC$GENEQ)
summ(lm2015G_SC)
ggplotRegression(lm2015G_SC)
#
lm2015D_SC <- lm(CORR2015_SC$PISA ~ CORR2015_SC$DEVEL)
summ(lm2015D_SC)
ggplotRegression(lm2015D_SC)
#
#-----
# Contingency tables
#-----
#
CONT2015_SC <- DATA_SC[which(DATA_SC$YEAR=="2015"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2015_SC$HIGENEQ, CONT2015_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2015_SC$HIGENEQ, CONT2015_SC$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2015_SC$HIDEV, CONT2015_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2015_SC$HIDEV, CONT2015_SC$LOWPIISA), margin=2)*100
#
#-----
# Subset data for correlations - 2018 - N=62
#-----
#
CORR2018_SC <- DATA_SC[which(DATA_SC$YEAR=="2018"),c("PISA", "GENEQ", "DEVEL")]
#
summary(CORR2018_SC$PISA)
sd(CORR2018_SC$PISA)
#
summary(CORR2018_SC$GENEQ)
sd(CORR2018_SC$GENEQ)
#
summary(CORR2018_SC$DEVEL)
sd(CORR2018_SC$DEVEL)
#
rcorr(as.matrix(CORR2018_SC))
#
# Quick visualization of relationships
lm2018G_SC <- lm(CORR2018_SC$PISA ~ CORR2018_SC$GENEQ)
summ(lm2018G_SC)
ggplotRegression(lm2018G_SC)
#
lm2018D_SC <- lm(CORR2018_SC$PISA ~ CORR2018_SC$DEVEL)
summ(lm2018D_SC)
ggplotRegression(lm2018D_SC)
#
#-----
# Contingency tables
#-----
#
CONT2018_SC <- DATA_SC[which(DATA_SC$YEAR=="2018"),c("HIGHPIISA", "LOWPIISA",
"HIGENEQ", "HIDEV")]
#
# Contingency table High PISA, HIGENEQ
prop.table(table(CONT2018_SC$HIGENEQ, CONT2018_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIGENEQ
prop.table(table(CONT2018_SC$HIGENEQ, CONT2018_SC$LOWPIISA), margin=2)*100
# Contingency table High PISA, HIDEV
prop.table(table(CONT2018_SC$HIDEV, CONT2018_SC$HIGHPIISA), margin=2)*100
# Contingency table Low PISA, HIDEV
prop.table(table(CONT2018_SC$HIDEV, CONT2018_SC$LOWPIISA), margin=2)*100
#

```



```
#####
# B) The relationship between change in gender equality and human
#   development, and change in PISA scores, 2006-2018.
#####
#
# Load new data set
TIMEDATA <- read_excel("PISA_DATA_TIME.xlsx")
#
head(TIMEDATA)
#
#-----
# Description of variables:
#-----
#
# P06-P18: PISA scores for 2006 - 2018
# GEN05-GEN18: (inverse of) GII scores for 2005 - 2018 (note, GII exists for 2005,
#             2010, and then every year, so 2005 is used for analyses with
#             PISA 2006)
# DEV06-DEC18: HDI scores for 2006 - 2018
# UNSTATE: UN STATE CLASSIFICATIONS (AFRICAN, ASIA-PACIFIC, EASTERN-EUR, LATIN-AC,
#             WESTERN-EUR)
# WBCLASS06, WBCLASS18: WORLD BAND ANALYTIC CLASSIFICATIONS: LOW, LOW-MID, UPPER-MID, HIGH
#
#####
# Make separate dataframes for average, reading, math, science
#####
#
TIMEDATA_AV <- TIMEDATA[which(TIMEDATA$SUBJECT=="Average"),]
TIMEDATA_RE <- TIMEDATA[which(TIMEDATA$SUBJECT=="Reading"),]
TIMEDATA_MA <- TIMEDATA[which(TIMEDATA$SUBJECT=="Math"),]
TIMEDATA_SC <- TIMEDATA[which(TIMEDATA$SUBJECT=="Science"),]
#
#####
# 1. Conduct analyses on AVERAGE
#####
#
#-----
# MAKE DATFRAME FOR CHANGE (12 YEARS) - POINT CHANGE
#-----
#
PISA_CHANGE_AV <- TIMEDATA_AV[,c("COUNTRY", "P06", "P18", "GEN05", "GEN18",
                                "DEV06", "DEV18", "UNSTATE", "WBCLASS06", "WBCLASS18")]
#
# Make variable for change between 2006 and 2018 for PISA, GENEQ, and DEVELOP
PISA_CHANGE_AV$P_CHANGE <- (PISA_CHANGE_AV$P18-PISA_CHANGE_AV$P06)
PISA_CHANGE_AV$G_CHANGE <- (PISA_CHANGE_AV$GEN18-PISA_CHANGE_AV$GEN05)
PISA_CHANGE_AV$D_CHANGE <- (PISA_CHANGE_AV$DEV18-PISA_CHANGE_AV$DEV06)
#
# OMIT ROWS WITH MISSING DATA
#
PISA_CHANGE_AV <- na.omit(PISA_CHANGE_AV)
#
# MOVE CHARACTER VARIABLES TO END OF FRAME
head(PISA_CHANGE_AV)
PISA_CHANGE_AV <- PISA_CHANGE_AV[,c("COUNTRY", "P06", "P18", "GEN05", "GEN18",
                                    "DEV06", "DEV18", "P_CHANGE", "G_CHANGE", "D_CHANGE",
                                    "UNSTATE", "WBCLASS06", "WBCLASS18")]
head(PISA_CHANGE_AV)
#
# Examine correlation matrix
rcorr(as.matrix(PISA_CHANGE_AV[2:7]))
summary(PISA_CHANGE_AV)
sd(PISA_CHANGE_AV$DEV18)
sd(PISA_CHANGE_AV$GEN18)
sd(PISA_CHANGE_AV$P06)
sd(PISA_CHANGE_AV$P18)
sd(PISA_CHANGE_AV$P_CHANGE)
sd(PISA_CHANGE_AV$G_CHANGE)
sd(PISA_CHANGE_AV$D_CHANGE)
#
# MODEL RELATIONSHIP OF P06 WITH P_CHANGE
lmP06_AV <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06)
summ(lmP06_AV)
#
#-----
# Model relationships change in GENDER
```

```

#-----
#
summary(PISA_CHANGE_AV$GEN05)
summary(PISA_CHANGE_AV$G_CHANGE)
#
# Regress P_CHANGE on G_CHANGE
lmGENDER_AV <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$G_CHANGE)
summ(lmGENDER_AV)
#
# Regress P_CHANGE on G_CHANGE, controlling for P06 and GEN05
lmGENDERcon_AV <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$G_CHANGE +
  PISA_CHANGE_AV$GEN05 + PISA_CHANGE_AV$P06)
summ(lmGENDERcon_AV)
#
# Check assumptions
plot(lmGENDERcon_AV,1)
plot(lmGENDERcon_AV,2)
plot(lmGENDERcon_AV,3)
plot(lmGENDERcon_AV,4)
plot(lmGENDERcon_AV,5)
hist(lmGENDERcon_AV$residuals,
  freq=FALSE,
  breaks=10,
  main="Histogram of residuals",
  xlab="Residuals",
  xlim=c(-50, 50),
  col="lightgrey")
lines(seq(-50, 50, by=.01),
  dnorm(seq(-50, 50, by=.01),
    mean(lmGENDERcon_AV$residuals),
    sd(lmGENDERcon_AV$residuals)), col="red")
#
#-----
# Model relationships change in HUMAN DEVELOPMENT
#-----
#
summary(PISA_CHANGE_AV$DEV06)
summary(PISA_CHANGE_AV$D_CHANGE)
#
# Regress P_CHANGE on D_CHANGE
lmDEV_AV <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$D_CHANGE)
summ(lmDEV_AV)
#
# Regress P_CHANGE on D_CHANGE, controlling for P06 and DEV06
lmDEVcon_AV <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$D_CHANGE +
  PISA_CHANGE_AV$DEV06 + PISA_CHANGE_AV$P06)
summ(lmDEVcon_AV)
#
#-----
# Evaluate controls for UN state and WB wealth categories
#-----
#
# MAKE SCATTER PLOT BY UNSTATE
ggplot(data = PISA_CHANGE_AV, aes(x = P06, y = P18, color=UNSTATE))+
  geom_point(lwd = 2) + xlab("PISA 2006 ") +
  ylab("PISA 2018")+ ylim(380, 555)+
  geom_abline(aes(intercept = 0, slope = 1), color = "grey", lwd = 1, lty = "longdash")+
  theme_classic2() + theme(legend.position = "right")
#
# MAKE SCATTER PLOT BY WBCLASS06
ggplot(data = PISA_CHANGE_AV, aes(x = P06, y = P18, color=WBCLASS06))+
  geom_point(lwd = 2) + xlab("PISA 2006 ") +
  ylab("PISA 2018")+ ylim(380, 555)+
  geom_abline(aes(intercept = 0, slope = 1), color = "grey", lwd = 1, lty = "longdash")+
  theme_classic2() + theme(legend.position = "right")
#
# MAKE SCATTER PLOT BY WBCLASS18
ggplot(data = PISA_CHANGE_AV, aes(x = P06, y = P18, color=WBCLASS18))+
  geom_point(lwd = 2) + xlab("PISA 2006 ") +
  ylab("PISA 2018")+ ylim(380, 555)+
  geom_abline(aes(intercept = 0, slope = 1), color = "grey", lwd = 1, lty = "longdash")+
  theme_classic2() + theme(legend.position = "right")
#
# Regress with dummy for UNSTATE
#
head(PISA_CHANGE_AV)

```

```

#
# Make dummy for UNSTATE 1=WESTERN EUROPE & OTHER
PISA_CHANGE_AV$WEST <- if_else(PISA_CHANGE_AV$UNSTATE=="WESTERN", 1, 0)
#
# Make dummy for UNSTATE 1=EASTERN EUROPE
PISA_CHANGE_AV$EAST <- if_else(PISA_CHANGE_AV$UNSTATE=="EASTEUR", 1, 0)
#
# Make dummy for UNSTATE 1=ASIA PACIFIC
PISA_CHANGE_AV$ASIA <- if_else(PISA_CHANGE_AV$UNSTATE=="ASIAPAC", 1, 0)
#
# Make dummy for UNSTATE 1=LATIN AMERICA & CARRIBEAN
PISA_CHANGE_AV$LATIN <- if_else(PISA_CHANGE_AV$UNSTATE=="LATINAC", 1, 0)
#
# Test all dummies together (reference category WEST)
lmDUMMY_UNALL <- lm(PISA_CHANGE_AV$P18 ~ PISA_CHANGE_AV$G_CHANGE + PISA_CHANGE_AV$P06 +
  PISA_CHANGE_AV$GEN05 + PISA_CHANGE_AV$LATIN + PISA_CHANGE_AV$ASIA +
  PISA_CHANGE_AV$EAST)
summ(lmDUMMY_UNALL)
#
# Regress with dummy for WBCLASS
#
head(PISA_CHANGE_AV)
#
# Make dummy for WBCLASS06 1=HIGH
PISA_CHANGE_AV$WBHIGH06 <- if_else(PISA_CHANGE_AV$WBCLASS06=="H", 1, 0)
#
# Make dummy for WBCLASS 1=MID
PISA_CHANGE_AV$WB MID06 <- if_else(PISA_CHANGE_AV$WBCLASS06=="H", 0, 1)
#
# Test (reference category WBHIGH06)
lmDUMMY_WB06 <- lm(PISA_CHANGE_AV$P18 ~ PISA_CHANGE_AV$G_CHANGE + PISA_CHANGE_AV$P06 +
  PISA_CHANGE_AV$GEN05 + PISA_CHANGE_AV$WB MID06)
summ(lmDUMMY_WB06)
#
# Make dummy for WBCLASS18 1=HIGH
PISA_CHANGE_AV$WBHIGH18 <- if_else(PISA_CHANGE_AV$WBCLASS18=="H", 1, 0)
#
# Make dummy for WBCLASS 1=MID
PISA_CHANGE_AV$WB MID18 <- if_else(PISA_CHANGE_AV$WBCLASS18=="H", 0, 1)
#
# Test (reference category WBHIGH18)
lmDUMMY_WB18 <- lm(PISA_CHANGE_AV$P18 ~ PISA_CHANGE_AV$G_CHANGE + PISA_CHANGE_AV$P06 +
  PISA_CHANGE_AV$GEN05 + PISA_CHANGE_AV$WB MID18)
summ(lmDUMMY_WB18)
#
#
# Build up models based on predicting P_CHANGE
#
Model0 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06)
summ(Model0)
#
# Test country effects (dummy variable WEST)
Model1_region <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$EAST +
  PISA_CHANGE_AV$LATIN + PISA_CHANGE_AV$ASIA)
summ(Model1_region)
#
# Test country income level effects 2006 (dummy variable WBHIGH06)
Model1_income06 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$WB MID06)
summ(Model1_income06)
#
# Test country income level effects 2018 (dummy variable WBHIGH18)
Model1_income18 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$WB MID18)
summ(Model1_income18)
#
#
# Test gender change
Model2 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$G_CHANGE)
summ(Model2)
#
# Test gender change with country effects (dummy WEST)
Model3_region <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$EAST +
  PISA_CHANGE_AV$ASIA + PISA_CHANGE_AV$LATIN +
  PISA_CHANGE_AV$G_CHANGE)
summ(Model3_region)
#
# Test gender change with country income effects 2006 (dummy WBHIGH06)

```

```

Model3_income06 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$WBMID06
+ PISA_CHANGE_AV$G_CHANGE)
summ(Model3_income06)
#
# Test gender change with country income effects 2018 (dummy WBHIGH18)
Model3_income18 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$WBMID18
+ PISA_CHANGE_AV$G_CHANGE)
summ(Model3_income18)
#
# Test gender change with all country effects (dummy WEST and WBHIGH18)
Model3_all <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$EAST +
PISA_CHANGE_AV$ASIA + PISA_CHANGE_AV$LATIN + PISA_CHANGE_AV$WBMID18 +
PISA_CHANGE_AV$G_CHANGE)
summ(Model3_all)
#
# Test human development
Model3_D <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$D_CHANGE)
summ(Model3_D)
#
# Test human development change with country effects (WEST as dummy)
Model4_D_region <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$EAST +
PISA_CHANGE_AV$ASIA + PISA_CHANGE_AV$LATIN +
PISA_CHANGE_AV$D_CHANGE)
summ(Model4_D_region)
#
# Test human development change with income country effects 2006 (WBHIGH06 as dummy)
Model4_D_income06 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$WBMID06 +
PISA_CHANGE_AV$D_CHANGE)
summ(Model4_D_income06)
#
# Test human development change with income country effects 2018 (WBHIGH18 as dummy)
Model4_D_income18 <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$WBMID18 +
PISA_CHANGE_AV$D_CHANGE)
summ(Model4_D_income18)
#
# Test human development change with all country effects (dummy WEST and WBHIGH18)
Model4_all <- lm(PISA_CHANGE_AV$P_CHANGE ~ PISA_CHANGE_AV$P06 + PISA_CHANGE_AV$EAST +
PISA_CHANGE_AV$ASIA + PISA_CHANGE_AV$LATIN + PISA_CHANGE_AV$WBMID18 +
PISA_CHANGE_AV$D_CHANGE)
summ(Model4_all)
#
# No significant country effects or D_CHANGE effect
#
#####
# Change in gender equality and change in PISA scores, 2006-2018, separately
# for boys and girls, and for reading, math, & science
#####
#
# Load new data set
TIMEDATABoygirl <- read_excel("PISA_2006_2018_boys_girls.xlsx")
#
head(TIMEDATABoygirl)
#
#-----
# MAKE DATFRAME FOR CHANGE (girls)
#-----
#
CHANGEgirl <- TIMEDATABoygirl[,c("COUNTRY", "SUBJECT", "P06_GIRLS", "P18_GIRLS",
"GEN05", "GEN18", "DEV06", "DEV18")]
#
# Make variable for change between 2006 and 2018 for PISA, GENEQ, and DEVELOP
CHANGEgirl$P_CHANGE <- (CHANGEgirl$P18_GIRLS-CHANGEgirl$P06_GIRLS)
CHANGEgirl$G_CHANGE <- (CHANGEgirl$GEN18-CHANGEgirl$GEN05)
CHANGEgirl$D_CHANGE <- (CHANGEgirl$DEV18-CHANGEgirl$DEV06)
#
# Make data sets for AVERAGE, READING, MATH, SCIENCE
CHANGEgirl_AV <- CHANGEgirl[which(CHANGEgirl$SUBJECT=="AVERAGE"),]
CHANGEgirl_RE <- CHANGEgirl[which(CHANGEgirl$SUBJECT=="READING"),]
CHANGEgirl_MA <- CHANGEgirl[which(CHANGEgirl$SUBJECT=="MATH"),]
CHANGEgirl_SC <- CHANGEgirl[which(CHANGEgirl$SUBJECT=="SCIENCE"),]
#
#-----
# MAKE DATFRAME FOR CHANGE (boys)
#-----
#
CHANGEboy <- TIMEDATABoygirl[,c("COUNTRY", "SUBJECT", "P06_BOYS", "P18_BOYS",

```

```

"GEN05", "GEN18", "DEV06", "DEV18"]])
#
# Make variable for change between 2006 and 2018 for PISA, GENEQ, and DEVELOP
CHANGEboy$P_CHANGE <- (CHANGEboy$P18_BOYS-CHANGEboy$P06_BOYS)
CHANGEboy$G_CHANGE <- (CHANGEboy$GEN18-CHANGEboy$GEN05)
CHANGEboy$D_CHANGE <- (CHANGEboy$DEV18-CHANGEboy$DEV06)
#
# Make data sets for AVERAGE, READING, MATH, SCIENCE
CHANGEboy_AV <- CHANGEboy[which(CHANGEboy$SUBJECT=="AVERAGE"),]
CHANGEboy_RE <- CHANGEboy[which(CHANGEboy$SUBJECT=="READING"),]
CHANGEboy_MA <- CHANGEboy[which(CHANGEboy$SUBJECT=="MATH"),]
CHANGEboy_SC <- CHANGEboy[which(CHANGEboy$SUBJECT=="SCIENCE"),]
#
#-----
# MAKE DATFRAME FOR CHANGE (all)
#-----
#
CHANGEall <- TIMEDATAboygirl[,c("COUNTRY", "SUBJECT", "P06_ALL", "P18_ALL",
                                "GEN05", "GEN18", "DEV06", "DEV18")]
#
# Make variable for change between 2006 and 2018 for PISA, GENEQ, and DEVELOP
CHANGEall$P_CHANGE <- (CHANGEall$P18_ALL-CHANGEall$P06_ALL)
CHANGEall$G_CHANGE <- (CHANGEall$GEN18-CHANGEall$GEN05)
CHANGEall$D_CHANGE <- (CHANGEall$DEV18-CHANGEall$DEV06)
#
# Make data sets for AVERAGE, READING, MATH, SCIENCE
CHANGEall_AV <- CHANGEall[which(CHANGEall$SUBJECT=="AVERAGE"),]
CHANGEall_RE <- CHANGEall[which(CHANGEall$SUBJECT=="READING"),]
CHANGEall_MA <- CHANGEall[which(CHANGEall$SUBJECT=="MATH"),]
CHANGEall_SC <- CHANGEall[which(CHANGEall$SUBJECT=="SCIENCE"),]
#
#-----
# Model and compare relationships , by groups
#-----
#
# AVERAGE score, change predicting change model GENDER
#
# ALL
lmAVERAGE_ALL <- lm(CHANGEall_AV$P_CHANGE ~ CHANGEall_AV$G_CHANGE + CHANGEall_AV$GEN05 +
                     CHANGEall_AV$P06_ALL)
summ(lmAVERAGE_ALL)
#
# GIRLS
lmAVERAGE_GIRLS <- lm(CHANGEgirl_AV$P_CHANGE ~ CHANGEgirl_AV$G_CHANGE + CHANGEgirl_AV$GEN05 +
                      CHANGEgirl_AV$P06_GIRLS)
summ(lmAVERAGE_GIRLS)
#
# BOYS
lmAVERAGE_BOYS <- lm(CHANGEboy_AV$P_CHANGE ~ CHANGEboy_AV$G_CHANGE + CHANGEboy_AV$GEN05 +
                     CHANGEboy_AV$P06_BOYS)
summ(lmAVERAGE_BOYS)
#
#
# READING score, change predicting change model GENDER
#
# ALL
lmREADING_ALL <- lm(CHANGEall_RE$P_CHANGE ~ CHANGEall_RE$G_CHANGE + CHANGEall_RE$GEN05 +
                   CHANGEall_RE$P06_ALL)
summ(lmREADING_ALL)
#
# GIRLS
lmREADING_GIRLS <- lm(CHANGEgirl_RE$P_CHANGE ~ CHANGEgirl_RE$G_CHANGE + CHANGEgirl_RE$GEN05 +
                     CHANGEgirl_RE$P06_GIRLS)
summ(lmREADING_GIRLS)
#
# BOYS
lmREADING_BOYS <- lm(CHANGEboy_RE$P_CHANGE ~ CHANGEboy_RE$G_CHANGE + CHANGEboy_RE$GEN05 +
                     CHANGEboy_RE$P06_BOYS)
summ(lmREADING_BOYS)
#
#
# MATH score, change predicting change model GENDER
#
# ALL
lmMATH_ALL <- lm(CHANGEall_MA$P_CHANGE ~ CHANGEall_MA$G_CHANGE + CHANGEall_MA$GEN05 +

```

```

        CHANGEall_MA$P06_ALL)
summ(lmMATH_ALL)
#
# GIRLS
lmMATH_GIRLS <- lm(CHANGEgirl_MA$P_CHANGE ~ CHANGEgirl_MA$G_CHANGE + CHANGEgirl_MA$GEN05 +
        CHANGEgirl_MA$P06_GIRLS)
summ(lmMATH_GIRLS)
#
# BOYS
lmMATH_BOYS <- lm(CHANGEboy_MA$P_CHANGE ~ CHANGEboy_MA$G_CHANGE + CHANGEboy_MA$GEN05 +
        CHANGEboy_MA$P06_BOYS)
summ(lmMATH_BOYS)
#
#
# SCIENCE score, change predicting change model GENDER
#
# ALL
lmSCIENCE_ALL <- lm(CHANGEall_SC$P_CHANGE ~ CHANGEall_SC$G_CHANGE + CHANGEall_SC$GEN05 +
        CHANGEall_SC$P06_ALL)
summ(lmSCIENCE_ALL)
#
# GIRLS
lmSCIENCE_GIRLS <- lm(CHANGEgirl_SC$P_CHANGE ~ CHANGEgirl_SC$G_CHANGE + CHANGEgirl_SC$GEN05 +
        CHANGEgirl_SC$P06_GIRLS)
summ(lmSCIENCE_GIRLS)
#
# BOYS
lmSCIENCE_BOYS <- lm(CHANGEboy_SC$P_CHANGE ~ CHANGEboy_SC$G_CHANGE + CHANGEboy_SC$GEN05 +
        CHANGEboy_SC$P06_BOYS)
summ(lmSCIENCE_BOYS)
#
#
# DEVELOPMENT (for comparison) - should not be statistically significant
#
# ALL
lmDEV_ALL <- lm(CHANGEall_AV$P_CHANGE ~ CHANGEall_AV$D_CHANGE +
        CHANGEall_AV$P06_ALL + CHANGEall_AV$DEV06)
summ(lmDEV_ALL)
#
# GIRLS
lmDEV_GIRL <- lm(CHANGEgirl_AV$P_CHANGE ~ CHANGEgirl_AV$D_CHANGE +
        CHANGEgirl_AV$P06_GIRLS + CHANGEgirl_AV$DEV06)
summ(lmDEV_GIRL)
#
# BOYS
lmDEV_BOY <- lm(CHANGEboy_AV$P_CHANGE ~ CHANGEboy_AV$D_CHANGE +
        CHANGEboy_AV$P06_BOYS + CHANGEboy_AV$DEV06)
summ(lmDEV_BOY)
#
#####
# C) The relationship between the items of gender equality and human
#   development 2018, and PISA scores 2018.
#####
#
# Load data set
#
HDI_GII <- read_excel("PISA_HDI_GII_CHANGE.xlsx")
#
head(HDI_GII)
#
#####
# Make separate dataframes for average, reading, math, science
#####
#
HDI_GII_AV <- HDI_GII[which(HDI_GII$SUBJECT=="AVERAGE"),]
HDI_GII_RE <- HDI_GII[which(HDI_GII$SUBJECT=="READING"),]
HDI_GII_MA <- HDI_GII[which(HDI_GII$SUBJECT=="MATH"),]
HDI_GII_SC <- HDI_GII[which(HDI_GII$SUBJECT=="SCIENCE"),]
#
# Work with just subset AVERAGE
#
#-----
# Description of variables
#-----
#
# H1 Life expectancy at birth

```

```

# H2 Expected years of schooling for children starting school
# H3 Mean years of schooling for adults aged over 25
# H4 Gross national income per capita
# G1 Maternal mortality deaths per 100,000 live births
# G2 Adolescent births per 1,000 women aged 15-19
# G3 Percentage of parliament seats held by women
# G4 Percentage of women with some secondary education
# G5 Difference in labour force participation male-female
#
# Create log transformed H4, as income typically has a curvilinear relationship
# with social outcomes
#
HDI_GII_AV$logH4_06 = log(HDI_GII_AV$H4_06)
HDI_GII_AV$logH4_18 = log(HDI_GII_AV$H4_18)
#
# Subset data for correlation matrix 2018
HDI_GII_AV_2018 <- HDI_GII_AV[,c("H1_18", "H2_18", "H3_18", "H4_18", "logH4_18", "G1_18",
                                "G2_18", "G3_18", "G4_18", "G5_18", "P_18")]
summary(HDI_GII_AV_2018)
sd(HDI_GII_AV_2018$H1_18)
sd(HDI_GII_AV_2018$H2_18)
sd(HDI_GII_AV_2018$H3_18)
sd(HDI_GII_AV_2018$H4_18)
sd(HDI_GII_AV_2018$G1_18)
sd(HDI_GII_AV_2018$G2_18)
sd(HDI_GII_AV_2018$G3_18)
sd(HDI_GII_AV_2018$G4_18)
sd(HDI_GII_AV_2018$G5_18)
#
rcorr(as.matrix(HDI_GII_AV_2018))
#
#
#####
# 4) The relationship between change in the items of gender equality and
#   human development, and change in PISA scores, 2006-2018. (AVERAGE)
#####
#
head(HDI_GII_AV)
#
# MAKE DATFRAME FOR CHANGE (12 YEARS)
#
HDI_GII_AV_CHANGE <- HDI_GII_AV[,c("Country", "P_06", "P_18")]
#
HDI_GII_AV_CHANGE$P_CHANGE <- (HDI_GII_AV$P_18-HDI_GII_AV$P_06)
HDI_GII_AV_CHANGE$H1_CHANGE <- (HDI_GII_AV$H1_18-HDI_GII_AV$H1_06)
HDI_GII_AV_CHANGE$H2_CHANGE <- (HDI_GII_AV$H2_18-HDI_GII_AV$H2_06)
HDI_GII_AV_CHANGE$H3_CHANGE <- (HDI_GII_AV$H3_18-HDI_GII_AV$H3_06)
HDI_GII_AV_CHANGE$H4_CHANGE <- (HDI_GII_AV$H4_18-HDI_GII_AV$H4_06)
HDI_GII_AV_CHANGE$G1_CHANGE <- (HDI_GII_AV$G1_18-HDI_GII_AV$G1_05)
HDI_GII_AV_CHANGE$G2_CHANGE <- (HDI_GII_AV$G2_18-HDI_GII_AV$G2_05)
HDI_GII_AV_CHANGE$G3_CHANGE <- (HDI_GII_AV$G3_18-HDI_GII_AV$G3_05)
HDI_GII_AV_CHANGE$G4_CHANGE <- (HDI_GII_AV$G4_18-HDI_GII_AV$G4_05)
HDI_GII_AV_CHANGE$G5_CHANGE <- (HDI_GII_AV$G5_18-HDI_GII_AV$G5_05)
HDI_GII_AV_CHANGE$logH4_CHANGE <- (HDI_GII_AV$logH4_18-HDI_GII_AV$logH4_06)
#
summary(HDI_GII_AV_CHANGE)
rcorr(as.matrix(HDI_GII_AV_CHANGE[3:13]))
#
# Follow same logic as before, and test the relationship between change in item
# and change in PISA, including P06 as a control, for each item
#
#-----
# Model relationship between change in H1: LIFE EXPECTANCY and change in PISA
#-----
#
# Regress P_CHANGE on H1_CHANGE and P_06
lmH1 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$H1_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmH1)
#
# H1_CHANGE is not statistically significant, when controlling for P06
#
#-----
# Model relationship between change in H2: EXPECTED YEARS OF SCHOOLING and
# change in PISA
#-----
#

```

```

# Regress P_CHANGE on H2_CHANGE and P_06
lmH2 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$H2_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmH2)
#
# H2_CHANGE is not statistically significant, when controlling for P06
#
#-----
# Model relationship between change in H3: MEAN YEARS OF SCHOOLING ADULTS and
# change in PISA
#-----
#
# Regress P_CHANGE on H3_CHANGE and P_06
lmH3 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$H3_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmH3)
#
# H3_CHANGE is not statistically significant, when controlling for P06
#
#-----
# Model relationship between change in H4: GNI PER CAPITA and change in PISA
#-----
#
# Regress P_CHANGE on H4_CHANGE and P_06
lmH4 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$H4_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmH4)
#
# H4_CHANGE is not statistically significant, when controlling for P06
#
# Check with logH4
#
lmH4log <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$logH4_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmH4log)
#
# logH4_CHANGE is not statistically significant, when controlling for P06
#
#-----
# Model relationship between change in G1: MATERNAL MORTALITY and change in PISA
#-----
#
# Regress P_CHANGE on G1_CHANGE and P_06
lmG1 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$G1_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmG1)
#
# G1_CHANGE is not statistically significant, when controlling for P06
#
#-----
# Model relationship between change in G2: ADOLESCENT BIRTHS and change in PISA
#-----
#
# Regress P_CHANGE on G2_CHANGE and P_06
lmG2 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$G2_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmG2)
#
#
# Graph of observed values
ggplot(lmG2, aes(HDI_GII_AV_CHANGE$G2_CHANGE, HDI_GII_AV_CHANGE$P_CHANGE)) +
  geom_point() +
  geom_hline(yintercept=0, color = "grey") +
  geom_vline(xintercept=0, color = "grey") +
  ggtitle("Relationship between change in adolescent birth rate \nand change in PISA score, \ncontrolling for PISA 2006")+
  xlab("Change in adolescent births per 1,000 women (15-19) 2005-2018")+
  ylab("Change in PISA score 2006-2018")
#
#
#-----
# Model relationship between change in G3: PERCENTAGE PARLIAMENT SEATS WOMEN and
# change in PISA
#-----
#
# Regress P_CHANGE on G3_CHANGE and P_06
lmG3 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$G3_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmG3)
#
# All variables, and final model, are statistically significant, p<0.001
#
#
# Graph observed relationship

```



```

ggplot(lmG3, aes(HDI_GII_AV_CHANGE$G3_CHANGE, HDI_GII_AV_CHANGE$P_CHANGE)) +
  geom_point() +
  geom_hline(yintercept=0, color = "grey") +
  geom_vline(xintercept=0, color = "grey") +
  ggtitle("Relationship between change in women's seats in parliament \nand change in PISA score, \ncontrolling for PISA 2006")+
  xlab("Change in percentage seats in parliament occupied by women 2005-2018")+
  ylab("Change in PISA score 2006-2018")
#
#
#-----
# Model relationships change in G4: PERCENTAGE OF WOMEN WITH SECONDARY ED.
#-----
#
# Regress P_CHANGE on G4_CHANGE and P_06
lmG4 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$G4_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmG4)
#
# G4_CHANGE is not statistically significant, when controlling for P06
#
#-----
# Model relationships change in G5: DIFFERENCE IN GENDER LABOUR FORCE PARTICIPATION
#-----
#
# Regress P_CHANGE on G5_CHANGE and P_06
lmG5 <- lm(HDI_GII_AV_CHANGE$P_CHANGE ~ HDI_GII_AV_CHANGE$G5_CHANGE + HDI_GII_AV_CHANGE$P_06)
summ(lmG5)
#
# G5_CHANGE is not statistically significant, when controlling for P06
#
#####
# GRAPHS AND FINAL MODELS
#####
#
#
#=====
# Make a graph for final gender model
#=====
#
lmAVERAGE_ALL_FINAL <- lm(CHANGEall_AV$P_CHANGE ~ CHANGEall_AV$G_CHANGE + CHANGEall_AV$GEN05 +
  CHANGEall_AV$P06_ALL )
summ(lmAVERAGE_ALL_FINAL)
#
#
summary(PISA_CHANGE_AV$P06)
summary(PISA_CHANGE_AV$GEN05)
summary(PISA_CHANGE_AV$G_CHANGE)
#
# Create simulation data
#
SIM_DATA <- data.frame (COUNTRY = c("A", "B", "C", "D", "E", "F", "G", "H", "I"),
  P06 = c(420, 420, 420, 480, 480, 480, 540, 540, 540),
  GEN05 = c(0.775, 0.775, 0.775, 0.775, 0.775, 0.775, 0.775, 0.775, 0.775),
  G_CHANGE = c(0.03, 0.06, 0.09, 0.03, 0.06, 0.09, 0.03, 0.06, 0.09))
SIM_DATA$P18_pred <- SIM_DATA$P06+90.09+(SIM_DATA$P06*-0.31)+(SIM_DATA$GEN05*56.76)+(SIM_DATA$G_CHANGE*178.71)
head(SIM_DATA)
#
# Make long version of the data
SIM_LONG <- SIM_DATA[,c("COUNTRY", "P06", "P18_pred", "G_CHANGE")]
SIM_LONG$G_CHANGE <- factor(SIM_LONG$G_CHANGE)
SIM_LONG <- gather(SIM_LONG, condition, measurement, P06:P18_pred, factor_key = T)
SIM_LONG
SIM_LONG$condition <- ifelse(SIM_LONG$condition=="P06", 0, 1)
SIM_LONG
#
#
ggplot(data = SIM_LONG, aes(x = condition, y = measurement, group = COUNTRY, color = G_CHANGE)) +
  geom_line(lwd = 1)+
  xlab("")+
  ylab("PISA score")+
  scale_y_continuous(breaks=c(420, 440, 460, 480, 500, 520, 540))+
  scale_x_continuous(breaks=c(0,1), labels=c("2006", "predicted \n2018"))+
  scale_color_grey(name = "Improvement in gender equality", labels = c("low", "mean", "high"))+
  theme_bw()+
  theme(legend.position="bottom")
#
#

```

```

# Graph change with colour for starting GENEQ value (high/low) and shape for
# G_CHANGE (high/low)
#
summary(PISA_CHANGE_AV$GEN05)
summary(PISA_CHANGE_AV$G_CHANGE)
# Make dummy for HIGHGEN05 (2005 OECD average for crossover)
PISA_CHANGE_AV$HIGHGEN05 <- if_else(PISA_CHANGE_AV$GEN05>=0.834, 1, 0)
PISA_CHANGE_AV$HIGHGEN05 <- as.factor(PISA_CHANGE_AV$HIGHGEN05)
# Make dummy for HIGHGENCH
PISA_CHANGE_AV$HIGHGENCH <- if_else(PISA_CHANGE_AV$G_CHANGE>=0.0618, 1, 0)
PISA_CHANGE_AV$HIGHGENCH <- as.factor(PISA_CHANGE_AV$HIGHGENCH)
#
# MAKE SCATTER PLOT BY HIGHGEN05
ggplot(data = PISA_CHANGE_AV, aes(x = P06, y = P18, color=HIGHGEN05))+
  geom_point(lwd = 2) + xlab("PISA 2006 ") +
  ylab("PISA 2018")+ ylim(380, 555)+
  geom_abline(aes(intercept = 0, slope = 1), color = "grey", lwd = 1, lty = "longdash")+
  theme_classic2() + theme(legend.position = "right")
#
# MAKE SCATTER PLOT BY HIGHGENCH
ggplot(data = PISA_CHANGE_AV, aes(x = P06, y = P18, color=HIGHGENCH))+
  geom_point(lwd = 2) + xlab("PISA 2006 ") +
  ylab("PISA 2018")+ ylim(380, 555)+
  geom_abline(aes(intercept = 0, slope = 1), color = "grey", lwd = 1, lty = "longdash")+
  theme_classic2() + theme(legend.position = "right")
#
#
# Simple scatter plot with mid line
#
ggplot(data = PISA_CHANGE_AV, aes(x = P06, y = P18))+
  geom_point(lwd = 2) + xlab("PISA 2006 ") +
  ylab("PISA 2018")+ ylim(380, 555)+
  geom_abline(aes(intercept = 0, slope = 1), color = "grey", lwd = 0.5, lty = "longdash")+
  theme_bw()
#
#
#-----
# Save Script
#-----
#
# *****Manually save script*****
#
#-----
# Quit R
#-----
q()

```