Dataset: 20200306 Data International Women’s Day

**Question 1**: With which statement did people agree and disagreed the most for which it is justified to hit or beat the wife?

**Response**: the most agreed statement was: “for at least one specific reason”, the least agreed statement was: “if she burns food” (**Fig. 1**).

**Question 2**: Which gender was more likely to agree that it was justified to hit or beat the wife?

**Response**: Women always significantly agreed more than men that it was justified to hit or beat a wife for any statement (**Fig. 1**)

Chart, box and whisker chart

Description automatically generated

**Fig. 1** Difference in value (i.e., the percentage of people agreeing that it was justified to hit or beat a wife) per statement (i.e., question) and per gender (F = female, M = male). Significant differences are illustrated by non-overlapping notches.

**Code used to produce plot in R**:

# to set your work directory to the folder containing the file:

setwd("folder location")

#load your dataframe

dat <- read.csv( "20200306 Data International Women's Day Viz5 Launch.csv")

# check in dataframe did load properly

head(dat)

#to get more info on the dataframe

str(dat)

#to upload package (install first if needed)

library(ggplot2)

#build the boxplot

ggplot(dat, aes(x = Question, y = Value))+

geom\_boxplot(aes(fill = Gender),notch=TRUE,las=2)+

theme(axis.text.x = element\_text(angle = 90))

**Question 3**: Where are the countries located that agreed and disagreed the most wetter it is justified or not to hit or beat a wife “for at least one specific reason”?

**Answer**: South America had the lowest values (i.e., disagreed the most) whereas countries in Africa and Asia the closest to the equator had the highest values (i.e., agreed the most). We were wondering why only countries in these regions took part in the survey.

Diagram

Description automatically generated

**Fig. 2** World map showing the value (i.e., the percentage of people agreeing that it was justified to hit or beat a wife) for each country that took part in the survey. Countries in grey did not take part in the survey.

**Code for map in R** (for setting work directory and loading dataframe, see code above):

#load package

library(rworldmap)

#subsample the dataframe with only entries for the question “…for at least one specific reason”

datq <- subset(dat,dat$Question=="... for at least one specific reason")

#build map

sPDF <- joinCountryData2Map(datq, joinCode='NAME', nameJoinColumn='Country')

mapCountryData (sPDF, nameColumnToPlot='Value',missingCountryCol='darkgrey',mapTitle ="... for at least one specific reason")

**Question 4**: Is the percentage of people agreeing that it was justified to hit or beat a wife correlated with the Gross Domestic Product (GDP) of these countries?

**Answer**: There is a slight correlation (Kendall's rank correlation tau, tau = -0.104999, z = -15.676, p-value < 2.2e-16) with the percentage of people agreeing that it was justified to hit or beat a wife correlated and the GDP of these countries, with countries having the highest share of world GDP to disagree the most with all statements (**Fig. 3**).

Chart

Description automatically generated

**Fig. 4** Scatter plot of the value (i.e., the percentage of people agreeing that it was justified to hit or beat a wife) versus the share of world GDP (in percentage) of the countries that participated in the survey.

**Code in R for statistics and plot:**

# we first added an additional column in the dataframe with the share of World GDP we found online (<https://www.worldometers.info/gdp/gdp-by-country/>)

# check if data is normally distributed to know which correlation test to use

shapiro.test(dat$Value)

#the data “Value” was not normally distributed, so we applied a non-parametric correlation test

cor.test(dat$GDP, dat$Value, method="kendall")

#built the plot

library(ggplot2)

plot(x=dat$GDP,y=dat$Value)

ggplot(dat, aes(x=GDP, y=Value))+

geom\_point()+

geom\_smooth(method=lm)