

3.4: Examples with Data

In the lab for correlation you will be shown how to compute correlations in real data-sets using software. To give you a brief preview, let's look at some data from the world happiness report (2018). This report measured various attitudes across people from different countries. For example, one question asked about how much freedom people thought they had to make life choices. Another question asked how confident people were in their national government. Here is a scatterplot showing the relationship between these two measures. Each dot represents means for different countries.

```
library(data.table)
library(ggplot2)
suppressPackageStartupMessages(library(dplyr))
whr_data <- fread("https://stats.libretexts.org/@api/deki/files/10477/WHR2018.csv")
# select DVs and filter for NAs
smaller_df <- whr_data %>%
  dplyr::select(country,
    `Freedom to make life choices`,
    `Confidence in national government`) %>%
  dplyr::filter(!is.na(`Freedom to make life choices`),
    !is.na(`Confidence in national government`))
# plot the data with best fit line
ggplot(smaller_df, aes(x=`Freedom to make life choices`,
  y=`Confidence in national government`))+
  geom_point(alpha=.5)+
  geom_smooth(method=lm, se=FALSE, formula=y ~ x)+
  theme_classic()
```

run

restart

restart & run all

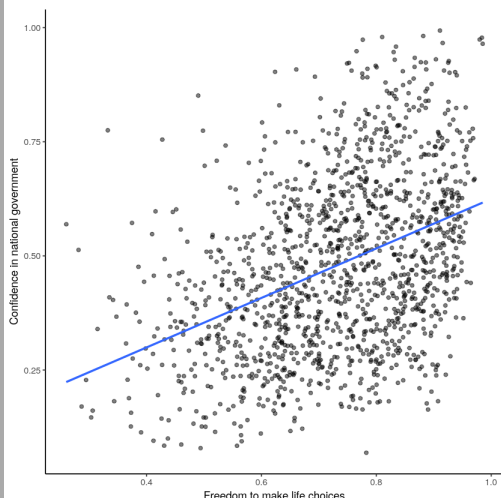


Figure 3.4.1: Relationship between freedom to make life choices and confidence in national government. Data from the world happiness report for 2018.

We put a blue line on the scatterplot to summarize the positive relationship. It appears that as “freedom to make life choices goes up”, so to does confidence in national government. It’s a positive correlation.

The actual correlation, as measured by Pearson’s r is:

```
library(data.table)
suppressPackageStartupMessages(library(dplyr))
whr_data <- fread("https://stats.libretexts.org/@api/deki/files/10477/WHR2018.csv")
# select DVs and filter for NAs
```

```
smaller_df <- whr_data %>%
  dplyr::select(country,
    `Freedom to make life choices`,
    `Confidence in national government`) %>%
  dplyr::filter(!is.na(`Freedom to make life choices`),
    !is.na(`Confidence in national government`))
# calculate correlation
cor(smaller_df$`Freedom to make life choices`,
  smaller_df$`Confidence in national government`)
```

run

restart

restart & run all

0.408096292505333

You will do a lot more of this kind of thing in the lab. Looking at the graph you might start to wonder: Does freedom to make life choices cause changes how confident people are in their national government? Or does it work the other way? Does being confident in your national government give you a greater sense of freedom to make life choices? Or, is this just a random relationship that doesn't mean anything? All good questions. These data do not provide the answers, they just suggest a possible relationship.

This page titled 3.4: Examples with Data is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Matthew J. C. Crump via source content that was edited to the style and standards of the LibreTexts platform.