

# Lab1

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2     3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr       1.0.2

-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(ggrepel)
data <- read_csv(file = "https://info3370.github.io/data/jencks_table1.csv")
```

```
Rows: 14 Columns: 4
-- Column specification -----
Delimiter: ","
chr (1): country
dbl (3): ratio, gdp, life_expectancy

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

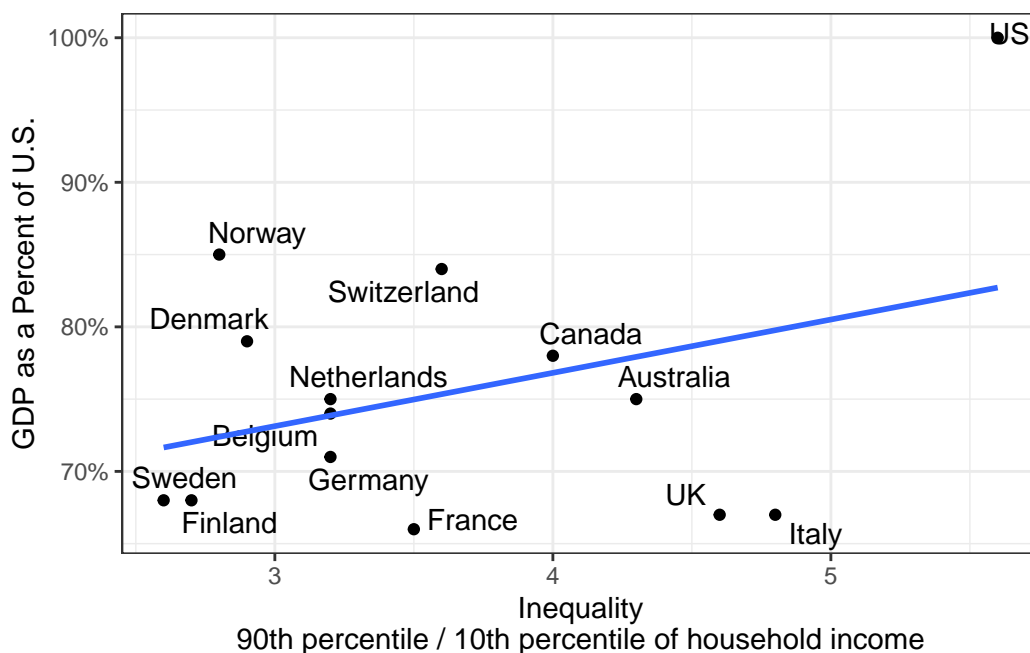
```
data |>
  ggplot(mapping = aes(x = ratio, y = gdp)) +
  geom_point() +
```

```

scale_x_continuous(
  name = "Inequality\n90th percentile / 10th percentile of household income"
) +
scale_y_continuous(
  name = "GDP as a Percent of U.S.",
  labels = scales::percent
) +
geom_text_repel(aes(label = data$country)) +
geom_smooth(method = lm, se = FALSE) +
theme_bw()

```

`geom\_smooth()` using formula = 'y ~ x'



From this graph and the positive-sloping line of best fit, it appears that there is a positive correlation between inequality ratio and GDP as a percent of the United States. As the inequality ratio increases, so does estimated GDP as a percent of the United States. However, the U.S. appears to be an outlier, as it has an extremely high inequality and GDP compared with the rest of the countries. This could make it seem like there is a stronger positive correlation between inequality ratio and GDP than there really is.

Questions:

- What would happen if we removed the U.S. from the data?
- How would it vary if we grouped the countries into regions? By geography?