# Lab 2

Leah Peterson

2024-02-20

## Packages

```r
library(rio)
library(here)
library(dplyr)
library(ggplot2)
library(ggflags)
library(ggtext)
library(tidyr)
```

## Section 1

```r
require(countrycode)

pisa       <- import(here("data/pisa.csv"))
pisa$iso   <- tolower(codelist[codelist$country.name.en %in% pisa$Country,]$iso2c)
pisa$diff  <- pisa$Math - pisa$Reading

labels1 <- pisa %>% filter(diff > 0)
labels2 <- pisa %>% filter(diff < 0)

pisa
```

```
          Country Math Reading iso diff
1       Australia  480     500  au  -20
2           Chile  410     450  cl  -40
3         Denmark  520     500  dk   20
4          France  495     490  fr    5
5          Greece  440     450  gr  -10
6           Japan  535     505  jp   30
7          Mexico  410     420  mx  -10
8     Switzerland  510     490  ch   20
9          Turkey  435     450  tr  -15
10  United States  480     500  us  -20
```

```r
pisa %>% arrange(diff) %>%  mutate(Country = factor(Country, levels = unique(Country))) %>%
  ggplot(aes(x = diff, y = Country, country = iso)) +
  geom_segment(aes(xend = 0, yend = Country), color = "black") +  # Lollipop segments
  geom_flag() +  # Flags at the end of each segment
  geom_text(data = labels1, aes(label = Country, x = diff, y = Country),
            hjust = -0.25,
            vjust = 0.5) +
```

```
geom_text(data = labels2, aes(label = Country, x = diff, y = Country),
          hjust = 1.25,
          vjust = 0.5) +
scale_x_continuous(limits = c(-45, 45), breaks = seq(-45, 45, by = 10)) +
labs(title = '**The difference in mathematics and reading scores from PISA assessment**'
     ) +
theme(panel.background   = element_blank(),
      axis.line.x = element_line(color = "black", linewidth = 1),
      axis.ticks.length.x = unit(5, "mm"),
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      axis.text.y = element_blank(),
      axis.line.y = element_blank(),
      axis.ticks.y = element_blank(),
      plot.title = element_markdown()
      ) +
annotate("text", x = -45, y = 10, label = "
Reading scores are higher
    than Math scores", vjust = 1, hjust = 0, size = 3)+
 annotate("text", x = 25, y = 2, label = "
Math scores are higher
  than Reading scores", vjust = 1, hjust = 0, size = 3) +
annotate("segment", x = -25, xend = -45, y = 10.01, yend = 10.01,
         arrow = arrow(length = unit(0.3, "cm")), color = "black") +
annotate("segment", x = 25, xend = 45, y = 2.01, yend = 2.01,
         arrow = arrow(length = unit(0.3, "cm")), color = "black")
```

## The difference in mathematics and reading scores from PISA assessment
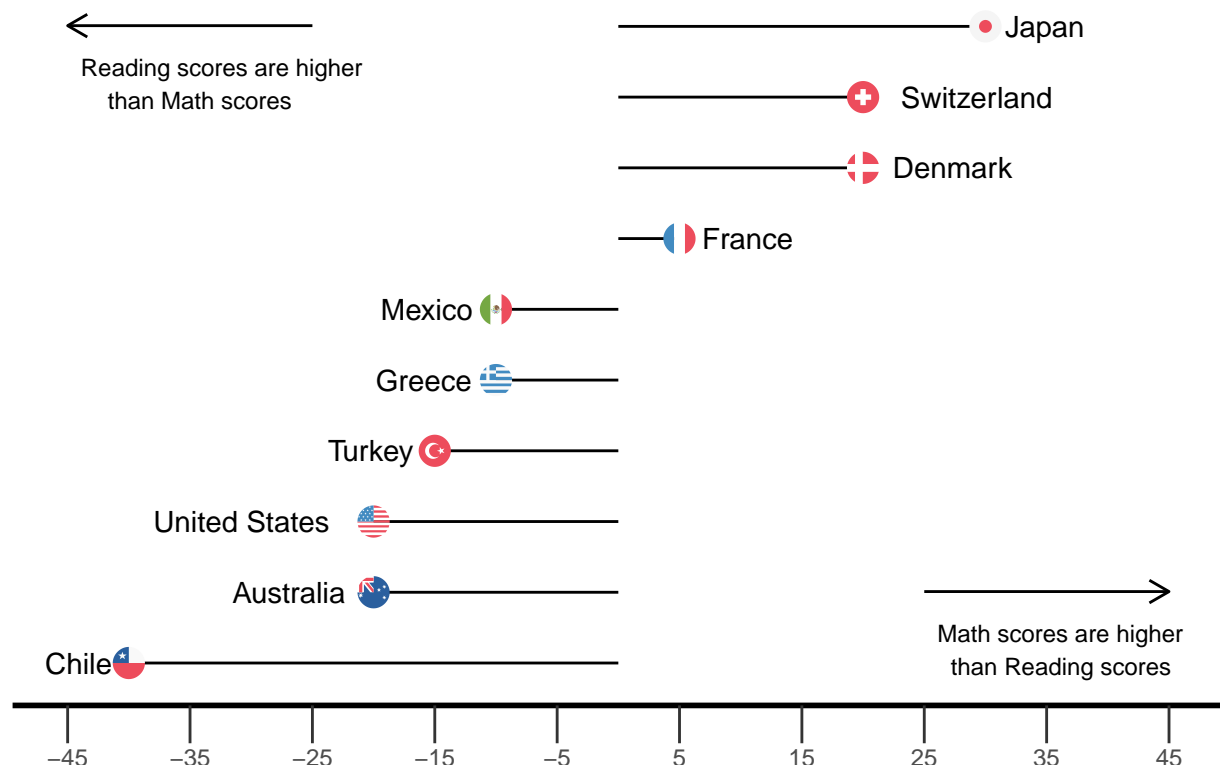
## Section 2

```
pop <- read.csv(here("data/population.csv"), fileEncoding = "UTF-8-BOM")

EU <- c("Austria", "Belgium", "Bulgaria", "Croatia", "Cyprus",
        "Czechia", "Denmark", "Estonia", "Finland", "France",
        "Germany", "Greece", "Hungary", "Ireland", "Italy", "Latvia",
        "Lithuania", "Luxembourg", "Malta", "Netherlands", "Poland",
        "Portugal", "Romania", "Slovak Republic", "Slovenia", "Spain","Sweden")

pop <- pop[pop$Country.Name %in% EU,]
pop <- pop[which(pop$X2022>10000000),]
pop$per_change <- ((pop$X2022 - pop$X1960)/pop$X2022)*100

pop_long <- pivot_longer(pop,
                         cols = starts_with("X"),
                         names_to = "Year",
                         values_to = "Population")

pop_long$Year <- as.integer(gsub("X", "", pop_long$Year))
my_colors <- c(Netherlands='orange',Spain='blue')
```

```
ggplot(pop_long, aes(x = Year, y = Population, color = Country.Name)) +
  geom_line(size = 1.5) +
  scale_color_manual(values = my_colors) +
  labs(title = "**<span style='color:blue'>Spain</span> and
      <span style='color:orange'>Netherlands</span>
       are the two countries with largest**  \n**population
      growth in European Union (1960-2022)**",
       subtitle = "(Among countries with at least 10 million people)  \n",
       y = "*Population  \n(in millions)*",
       caption = "Source: OECD Stats  \n https://stats.oecd.org/") +
  theme_minimal() +
 scale_y_continuous(limits = c(0, 85000000), breaks = seq(0, 80000000, by = 10000000),
                     labels = function(x) x / 1000000) +
  scale_x_continuous(breaks = seq(min(pop_long$Year), max(pop_long$Year), by = 20),
                     expand = c(0,0)) +
  theme(legend.position = "top",
        legend.title = element_blank(),
        panel.grid.major.y = element_line(linetype = "dashed"),
        panel.grid.major.x = element_blank(),
        panel.grid.minor = element_blank(),
        axis.title.x = element_blank(),
        axis.title.y = element_markdown(size = 10, vjust = 1.03, angle = 360,
                                        margin = margin(r = -200)),
        plot.title = element_markdown(),
        axis.line.x = element_line(color = "black", linewidth = .25),
        axis.ticks.x = element_line(color = "black", size = 0.25),
        plot.caption = element_text(hjust = 0)
        ) +
  guides(color = FALSE)
```

## **Spain** and **Netherlands** are the two countries with largest population growth in European Union (1960–2022)

(Among countries with at least 10 million people)

Population
(in millions)



Source: OECD Stats
https://stats.oecd.org/
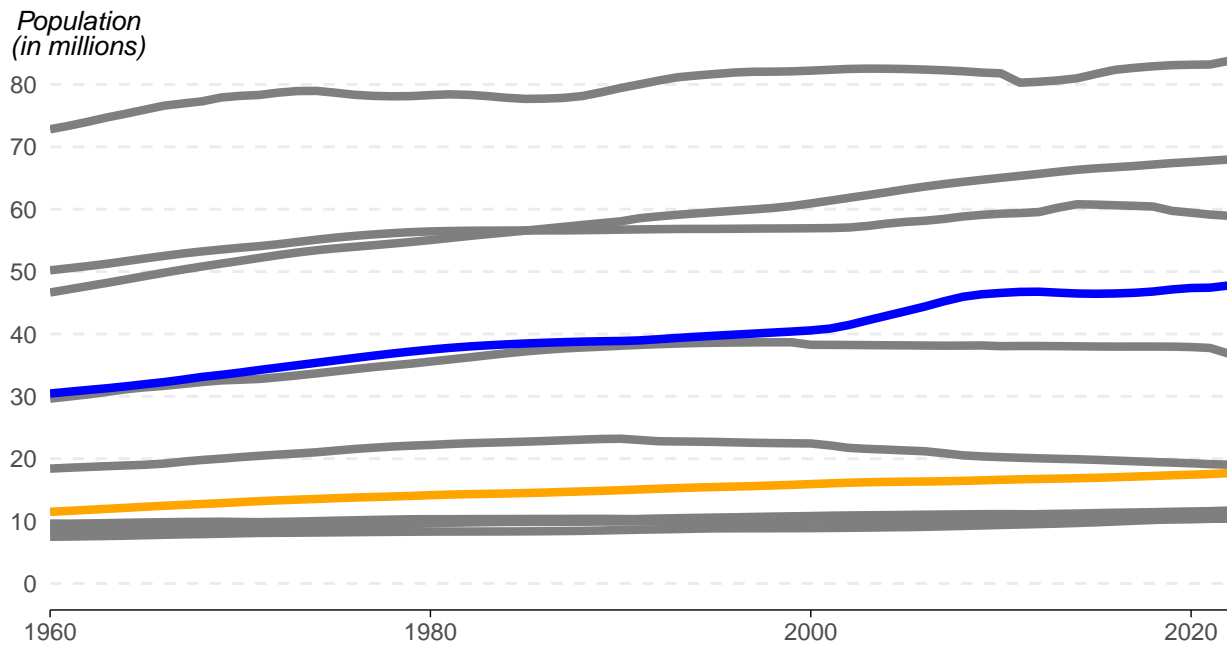
## Section 3

```r
hotel <- read.csv(here('data/hotel.csv'),fileEncoding="UTF-8-BOM")
hotel$travel_agents <- hotel$travel_agents/1000

  # number of travel agents are in thousands
  # hotel revenue is in billion dollars

hotel
```

```
   year travel_agents hotel_revenue
1  2000       123.385         12.95
2  2001       110.583         19.95
3  2002       104.046         28.02
4  2003       103.501         40.12
5  2004        90.428         51.16
6  2005        88.521         64.10
7  2006        87.431         79.81
8  2007        85.252         89.79
9  2008        86.070         94.46
10 2009        76.809         90.00
11 2010        70.272         99.76
12 2011        67.276        116.11
13 2012        64.552        124.60
14 2013        64.280        143.49
15 2014        63.975        155.38
```

```r
hotel1 <- hotel %>% filter(year >= 2000 & year <= 2004)
hotel2 <- hotel %>% filter(year >= 2004 & year <= 2008)
hotel3 <- hotel %>% filter(year >= 2008 & year <= 2009)
hotel4 <- hotel %>% filter(year >= 2009 & year <= 2014)



ggplot() +
    geom_path(data=hotel1,aes(x=hotel_revenue,y=travel_agents),color='#D95F02',
              linewidth=0.75) +
    geom_path(data=hotel2,aes(x=hotel_revenue,y=travel_agents),color='#0C72AF',
              linewidth=0.75) +
    geom_path(data=hotel3,aes(x=hotel_revenue,y=travel_agents),color='#00A174',
              linewidth=0.75) +
    geom_path(data=hotel4,aes(x=hotel_revenue,y=travel_agents),color='#D95F02',
              linewidth=0.75) +
    geom_point(data=hotel,aes(x=hotel_revenue,y=travel_agents),fill = "white",
              color = "black", size = 2, shape = 21) +
    geom_text(data = hotel, aes(x = hotel_revenue, y = travel_agents,
              label = year),
              vjust = -1.25,
              hjust = ifelse(hotel$year == 2009, 1, 0.1),
              size = 2) +
    scale_y_continuous(limits = c(0, 140), breaks = seq(0, 140, by = 20),
                       expand = expansion(mult = c(0, 0.05)),
                       labels = function(x) paste0(x, "K")) +
    scale_x_continuous(limits = c(0, 180), breaks = seq(0, 180, by = 30),
                       expand = c(0,1),
```

```r
                        labels = function(x) paste0("$", x, "B")) +
    geom_vline(xintercept = 0, color = "grey", linewidth = 0.75) +
    geom_hline(yintercept = 0, color = "grey", linewidth = 0.75) +
    theme_minimal() +
    labs(title = "**Online Hotel Revenue vs. Number of Travel Agents**",
         x = "Online Hotel Revenue (USD",
         y = "Number of Travel Agents") +
    theme(panel.grid.major = element_line(color = "grey", linetype = "dashed",
                              linewidth = 0.2),
         panel.grid.minor = element_blank(),
         panel.border = element_blank(),
         axis.line = element_blank(),
         plot.title = element_markdown(),
         axis.title.x = element_text(hjust = 0)
         ) +
  annotate("text", x = 22, y = 130,
  label = "Between 2000 and 2004, online hotel revenue increased
while the number of travel agents decreased",
  vjust = 1, hjust = 0, size = 2.5, color = '#D95F02') +
  annotate("text", x = 55, y = 105,
  label = "From 2004 to 2008, online hotel revenues continued to increase
while the number of travel agents decreased steadily",
  vjust = 1, hjust = 0, size = 2.5, color = '#0C72AF') +
  annotate("text", x = 105, y = 85,
  label = "Recession",
  vjust = 1, hjust = 0, size = 2.5, color = '#00A174') +
  annotate("segment", x = 95, xend = 104, y = 84, yend = 84,
           arrow = arrow(length = unit(0.1, "cm")), color = "black") +
  annotate("text", x = 30, y = 75,
  label = "From 2008 to 2009, online hotel revenue
  decreased a little bit for the first time
  since 2000 alongn with a sudden decrease
  in the travel agents",
  vjust = 1, hjust = 0, size = 2.5, color = '#00A174') +
  annotate("text", x = 115, y = 55,
  label = "Online hotel revenue recovered and continued
  to increase while the number of travel agents
  kept decreasing and stablilized",
  vjust = 1, hjust = 0, size = 2.5, color = '#D95F02')
```

# Online Hotel Revenue vs. Number of Travel Agents

Between 2000 and 2004, online hotel revenue increased while the number of travel agents decreased

From 2004 to 2008, online hotel revenues continued to increase while the number of travel agents decreased steadily

From 2008 to 2009, online hotel revenue decreased a little bit for the first time since 2000 alongn with a sudden decrease in the travel agents

Online hotel revenue recovered and continued to increase while the number of travel agents kept decreasing and stablilized

Recession

2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

**Number of Travel Agents**

**Online Hotel Revenue (USD**