Lab 4

Emmanuel Maduneme

2023-02-20

Getting Started

You can download the `transit_cost.csv` data from the website.

```
require(tidyverse)
require(lubridate)
require(ungeviz)
require(dviz.supp)
require(ggtext)
require(countrycode)
require(pacman)
p_load(here, rio, colorspace, scales, psych,
       flextable, transformr, purrr, ggridges,
       gganimate, ungeviz, dviz.supp, plotrix,
       colorblindr)


# remotes: :install_github ('clauswilke/dviz.supp')
# remotes: :install_github ('wilkelab/ungeviz')


here()
```

```
## [1] "/Users/emmanuelmaduneme/Desktop/UO SOJC PhD classes/Winter 23 Materials/Data Viz/Data Viz Proje
```

```
transit_cost <- read_csv(here("data", "transit_cost.csv"))
```

Question 1

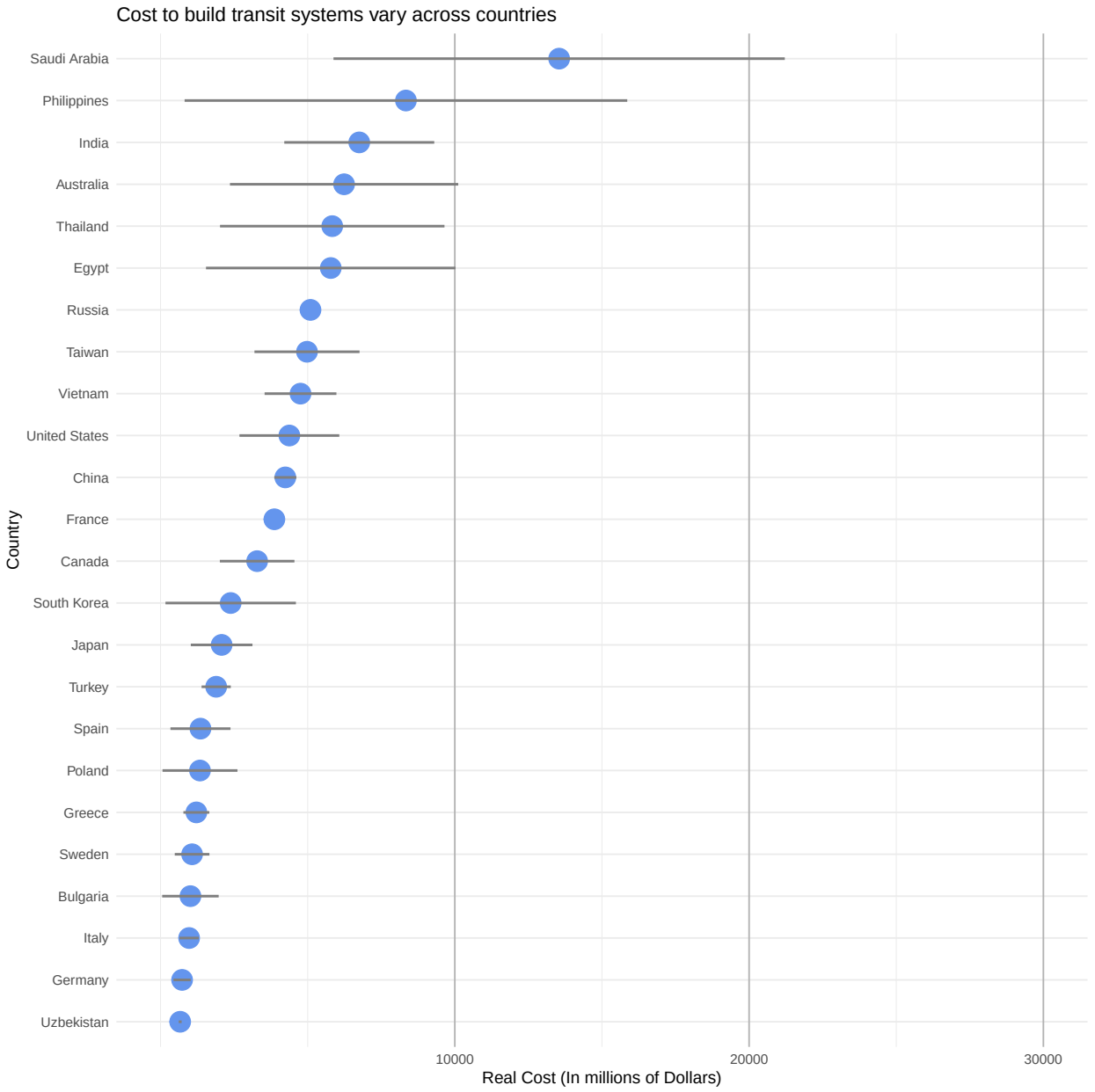
Use the transit costs data to reproduce the following plot. To do so, you will need to do a small amount of data cleaning, then calculate the means and standard errors (of the mean) for each country. Please filter to only counties with at least three observations. To use actual country names, rather than abbreviations, join your dataset with the output from the following

```
country_codes <- countrycode::codelist %>%
  select(country_name = country.name.en, country = ecb)

transit_country <- right_join(transit_cost, country_codes)
 transit_country <- transit_country %>%
  select(1, 21, 17) %>%
```

```
  group_by(country_name) %>%
  filter(n() > 3) %>%
  mutate(real_cost = as.numeric(real_cost)) %>%
   na.omit()

d <- describeBy(transit_country$real_cost, transit_country$country_name, mat=TRUE)[,c("group1","n", "me

d$low <- d$mean - 1.96*d$se
 d$high <- d$mean + 1.96*d$se

 d$group1 <- fct_reorder(factor(d$group1), d$mean)

 d %>%
   head() %>%
  flextable()
```

group1	n	mean	se	low	high
Australia	4	6,237.600	1,978.500	12,359.739	90 10,115.460
Bulgaria	6	1,016.295	489.4529	56.96729	1,975.623
Canada	10	3,282.997	646.5101	2,015.837	19 4,550.157
China	253	4,240.633	191.5129	3,865.267	54 4,615.998
Egypt	7	5,784.643	2,162.967	1,545.226	42 10,024.059
France	15	3,868.451	2,991.5348	-1,994.957	26 9,731.859

```
d %>%
  ggplot(aes(group1, mean))+
  geom_point(col = "cornflowerblue", size = 6) +
  coord_flip() +
  geom_linerange(aes(ymin= low,
                     ymax = high), width = 0.9, size = 0.8, color = "grey50") +
  scale_y_continuous(limits = c(0, 30000),
                     breaks = c(10000, 20000, 30000)) +
  theme_minimal() +
  theme(panel.grid.major.x = element_line(color = "grey70",
                                          size = 0.5,
                                          linetype = 1),
        plot.caption = element_text("Data provided through #tidytuesday by Transit Costs Project")) +
 labs(title = "Cost to build transit systems vary across countries",
      x = "Country",
      y = "Real Cost (In millions of Dollars)")
```



Question 2

A local news source reported on Nov 3, 2022, that the percentage of voters supporting Measure 114 was 46.1%. This estimate was based on only 577 voters; therefore, it has a margin of error of 4.1%.

Assume that the margin of error represents twice the standard error of the percentage estimate. Based on this information, create a quantile dot plot to represent the probability that Measure 114 would pass (more than 50% of all voters would support it).

Source: Oregon gun control Measure 114 polls closely

```
mu  <- .461
me <- 4.1  ## margin of error
se <- me/2  ## standard error
```

```r
# margin of error 4.1 = 1.96*SD
std <- round(me/1.96, digits = 1)    ## std = margin of error/1.96

x <- seq(0.01,.99,.01)

# Quantile data frame based on normal distribution

  m_114 <- data.frame(x = x,
                      q = qnorm(x, mu, std))
 ?qnorm
  m_114$winner <- ifelse(m_114$q <= .502,"#67a9cf","#ef8a62")


   ggplot(m_114, aes(q)) +
    geom_dotplot(aes(fill=winner),binwidth=0.24)+
    theme_minimal() +
     geom_vline(xintercept = 0,
                color = "gray",
                linetype = "dashed",
                size = 1)+
    scale_fill_identity(guide='none')+
    scale_y_continuous(name = "",
                       breaks = NULL) +
      ggtitle("Probability that Measure 114 would pass") +
  xlab("Proportion") +
  ylab("Probability") +
  theme_bw()
```

Probability that Measure 114 would pass