

```

---
title: "Problem Set 3"
name: "Clare Porter"
subtitle: "Due date: 30 September"
format:
  html:
    self-contained: true
toc: true
editor: visual
execute:
  warning: false
bibliography: references.bib
---

```

Please upload your completed assignment to the ELMs course site (under the assignments menu) and to your class Github repository. You need to upload both your Quarto document (with R code, as needed) and its rendered output (either as a PDF or HTML file).

****Total points: 27****

```

```{r}
dem_wealth_df <- readr::read_csv("http://svmiller.com/extdata/democracy-income-1950.csv")
|>
 dplyr::transmute(iso3c, country, gdp_per_cap = wbgdp2011est / wbpopest, democracy_score
= xm_qudsest)
```

```

The `dem_wealth_df` data set (taken from the wonderful Dr Steven V Miller's [blog post] (<https://svmiller.com/blog/2023/09/democracy-income-correlation-analysis/>)) will help us explore the relationship between a country's level of democracy and its citizens' wealth in 1950. It is a replication of [Lipset (1959)] (<https://www.cambridge.org/core/journals/american-political-science-review/article/abs/some-social-requisites-of-democracy-economic-development-and-political-legitimacy/26559429359F42D3E9B8BC82CA65546A>). Miller uses Xavier Marquez's [measure for democracy] (<https://www.cambridge.org/core/journals/political-analysis/article/democratic-compromise-a-latent-variable-analysis-of-ten-measures-of-regime-type/2A6B2BBA6F80367644F2C5007E1CFC29>). This is a great (if complicated) measure. For this problem set, you only need to know that a country's level of democracy increases with its "democracy score".

Question 1

Part A

Points: 2

> Using the `dem_wealth_df` data frame loaded in above, produce a scatter plot of the variables `democracy_score` and `gdp_per_cap` (with `gdp_per_cap` as the dependent variable on the y-axis). Professionally format your graph.

```

```{r}
ggplot(dem_wealth_df, aes(x = democracy_score, y = gdp_per_cap)) +
 geom_point() +
 theme_minimal() +
 labs(title = "Democracy and Citizen's Weath",
 x = "Democracy Score",
 y = "GDP per capita (USD current)") +
 scale_x_continuous()
```

```

Part B

Points: 3

> Describe the scatter plot. Note suspected outliers if any exist (a visual inspection will suffice for this question).

The scatter plot shows a relatively increasing relationship which could be fit to a regression. A low democracy score is associated with a low GDP per capita, and a higher democracy score is associated with a higher GDP per capita. Though not weak, the relationship doesn't appear to be very strong, with a few potential outliers being visibly apparent. Namely those data points with a GDP per capita over $\$15,000$. The data point which is just short of $\$10,000$ despite a democracy score below zero also seems like a plausible outlier.

Question 2

Points: 2

> Calculate the correlation coefficient between `democracy_score` and `gdp_per_cap`. What does it tell you about the relationship between a country's level of democracy and its citizens' wealth?

Using the code below, I find the correlation coefficient between democracy score and GDP per capita to be 0.7. A correlation coefficient of 0.7 tells us the relationship between our variables is not a perfect correlation but relatively strong and positive.

```
```{r}
cor(dem_wealth_df$democracy_score, dem_wealth_df$gdp_per_cap, use = "complete.obs")
```
```

Question 3

I would like to know whether individuals with higher levels of education tend to pay greater levels of attention to politics than those with lower levels of education.

Part A

Points: 5

> Using the `nes2024` data set in the `polisciols` R package, construct a cross-tab of the relationship between `attention_to_politics` and `education`.

```
```{r}
install.packages("devtools")
devtools::install_github("hgoers/polisciols")
library(modelsummary)
```

```
Save dataset
polisciols::nes2024
nes2024 <- polisciols::nes2024
```

```
datasummary_crosstab(attention_to_politics ~ education, data = nes2024)
```

## Let's make a bar graph like we did in class to better visualize:

```
nes2024 |>
 count(education, attention_to_politics) |>
 drop_na(education) |>
 ggplot(aes(x = n, y = education, fill = attention_to_politics)) +
 geom_bar(position = "dodge", stat = "identity") +
 theme_minimal() +
 labs(x = "N",
 y = NULL,
 fill = "Attention To Politics") +
 scale_fill_manual(values = c("red", "pink", "orange", "violet", "purple", "lightblue"))
```
```

Part B

Points: 3

> Describe the relationship between these two variables (if one exists).

The relationship between education level and attention to politics is unimodal and skewed right towards higher education and peaking in attention to politics amongst those completed some post-high school, but no bachelors degree. The less attention paid to politics, the less right skewed is the data towards higher education, with high school graduates paying the least attention to politics (never).

Question 4

Points: 5

I hypothesize that US voters who identify as being religious are less likely to support Federal government funding of scientific research. To test this, I run a survey of a representative sample of the US voting population. I ask respondents to indicate how often they attend religious ceremonies as a proxy measure of how strongly they identify as religious. They can answer: never, only for significant events, or regularly. Next, I ask respondents whether they think the Federal government's funding of scientific research is "too little", "about right", or "too much".

> Describe how you would construct a cross-tab to evaluate my hypothesis. Which responses (religious identification or support for Federal funding) would you compare within? Relatedly, which would you vary? Finally, what statistics would you use to make your comparisons?

I would compare religious identification (as the Independent Variable) with support for Federal Funding (as the Dependent Variable), varying religious identification. My code would have to first mutate the data to create categories of religious level and federal funding:

```
```{r}
US_voting_population <- US_voting_population |>
mutate(
federal_funding = case_when(
ntile(democracy_score, 3) == 1 ~ "Too Little",
ntile(democracy_score, 3) == 2 ~ "About Right",
ntile(democracy_score, 3) == 3 ~ "Too Much"
),
religious_ID = case_when(
ntile(gdp_per_cap, 3) == 1 ~ "Never",
ntile(gdp_per_cap, 3) == 2 ~ "Significant Events Only",
ntile(gdp_per_cap, 3) == 3 ~ "Regularly"
)
)
```
```

My code for my cross tab could then look something like:

```
```{r}

A cross-tab would look like:
datasummary_crosstab(religious_ID ~ federal_funding, data = US_voting_population)

A bar graph code would look like :
US_voting_population |>
count(education, religious_ID) |>
drop_na(education) |>
ggplot(aes(x = n, y = federal_funding, fill = religious_ID)) +
geom_bar(position = "dodge", stat = "identity") +
theme_minimal() +
labs(x = "N",
y = NULL,
```

```
fill = "Religious ID") +
scale_fill_manual(values = c("red","pink" "purple"))
```

```

This bar graph would reveal how various levels of federal funding is associated with religious identity level, and the shape of the relationship. Calculation of the correlation coefficient would provide further description of the direction and strength of this relationship between variables.

Question 5

Let's return to our original question of the relationship between democracy and individuals' wealth.

Part A

Points: 2

> Create two new variables based on a country's level of democracy, and its citizens' wealth. For the democracy variable, split countries into three categories: `autocracy` if they score in the lowest third of democracy scores, `hybrid` if they score in the middle third, and `democracy` if they score in the highest third. Similarly, for the wealth variable, split countries into three categories: `low income` if they score in the lowest third of GDP per capita, `middle income` if they score in the middle third, and `high income` if they score in the highest third.

```
```{r}
dem_wealth_df2 <- dem_wealth_df |>
 mutate(
 democracy_level = case_when(
 ntile(democracy_score, 3) == 1 ~ "Autocracy",
 ntile(democracy_score, 3) == 2 ~ "Hybrid",
 ntile(democracy_score, 3) == 3 ~ "Democracy"
),
 income_level = case_when(
 ntile(gdp_per_cap, 3) == 1 ~ "Low income",
 ntile(gdp_per_cap, 3) == 2 ~ "Middle income",
 ntile(gdp_per_cap, 3) == 3 ~ "High income"
)
)

```

```
dem_wealth_df2
```

```

Part B

Points: 3

> Create a cross-tab.

```
```{r}
Force into the correct order for display
dem_wealth_df2 <- dem_wealth_df2 |>
 mutate(
 democracy_level = factor(democracy_level,
 levels = c("Autocracy", "Hybrid", "Democracy")),
 income_level = factor(income_level,
 levels = c("Low income", "Middle income", "High income"))
)

```

## Then make cross-tab

```
datasummary_crosstab(democracy_level ~ income_level, data = dem_wealth_df2)
```

```

Part C

Points: 5

> Describe the relationship between these variables.

The relationship between democracy level and income level appears to be relatively strong and positive. As income level increases, countries are more likely to be democratic. Conversely, the lower income a country, the more likely the country is to be an autocracy or hybrid.

```
```{r}
Let's make another bar graph to better visualize the relationship:
dem_wealth_df2 |>
 count(income_level, democracy_level) |>
 ggplot(aes(x = n, y = income_level, fill = democracy_level)) +
 geom_bar(position = "dodge", stat = "identity") +
 theme_minimal() +
 labs(x = "N",
 y = NULL,
 fill = "Democracy Level") +
 scale_fill_manual(values = c("hotpink", "darkblue", "purple"))

```
```