Lab 2: R, Python, and Git, Oh My!

Due: 11:59 PM on 2/10/2025

Steven & Sriram

2025-02-06

```
import pandas as pd
```

Data Manipulation

Q1

```
# Reading in the data and then subsetting it to only keep data from the year 2019.

data <- read.csv("vdem_subset.csv")
data <- data[data$year == 2019, ]
```

```
# Same thing in python, except we can use the r version of the data.

pydata = r.data
pydata = pydata[pydata["year"] == 2019]
```

Q2

```
# Filter out NA values and then find the average GDP per capita. Use an ifelse to apply a Hig

data <- data[!is.na(data$e_gdppc), ]

average_gdp <- mean(data$e_gdppc)
data$gdp_cat <- ifelse(
  data$e_gdppc > average_gdp,
```

```
  "High",
  "Low"
)
```

```
# Same thing in python but use the .apply() method to apply a lambda function to the data tha

pydata = r.data
py_average_gdp = r.average_gdp # pydata["e_gdppc"].mean()

pydata["gdp_cat"] = pydata["e_gdppc"].apply(
  lambda x: "High" if x > py_average_gdp else "Low"
)
```

Q3

```
# Use the summarise function to find the average electoral democracy rating grouped by the hi

summary <- data %>%
  group_by(gdp_cat) %>%
  summarise(avg_electdem = mean(v2x_polyarchy), na.rm = TRUE)

print(summary)
```

```
# A tibble: 2 x 3
  gdp_cat avg_electdem na.rm
  <chr>          <dbl> <lgl>
1 High           0.659 TRUE
2 Low            0.452 TRUE
```

```
# Same thing in python but we can use the groupby method to group by and then stack a mean me

py_summary = pydata.groupby("gdp_cat")["v2x_polyarchy"].mean().reset_index(name="avg_electden
print(py_summary)
```

```
  gdp_cat  avg_electdem
0    High      0.659048
1     Low      0.452304
```

Q4

```
# Use a logical statement to filter countries with high GDP but low electoral democracy ratin

avg_electdem <- mean(data$v2x_polyarchy, na.rm = TRUE)
filtered_data <- data[
  data$e_gdppc > average_gdp & data$v2x_polyarchy < avg_electdem,
]
print(filtered_data)
```

```
            country_name year v2x_polyarchy e_gdppc e_regiongeo gdp_cat
34                Russia 2019         0.268  25.606           4    High
106          South Sudan 2019         0.174  21.725           8    High
350                Qatar 2019         0.081  80.190          10    High
370               Turkey 2019         0.289  23.599          10    High
402              Belarus 2019         0.257  20.806           4    High
434              Eritrea 2019         0.070  21.058           8    High
458           Kazakhstan 2019         0.237  23.827          11    High
518         Turkmenistan 2019         0.149  25.029          11    High
538              Bahrain 2019         0.118  30.046          10    High
582    Equatorial Guinea 2019         0.179  21.076           7    High
618               Kuwait 2019         0.315  45.010          10    High
634             Malaysia 2019         0.477  26.265          13    High
658                 Oman 2019         0.177  21.730          10    High
674         Saudi Arabia 2019         0.016  33.263          10    High
686            Singapore 2019         0.383  72.025          13    High
706 United Arab Emirates 2019         0.095  64.628          10    High
710              Hungary 2019         0.460  25.834           4    High
```

```
# Same thing in python

py_average_electdem = pydata["v2x_polyarchy"].mean()
py_filtered_data = pydata[
    (pydata["e_gdppc"] > py_average_gdp) & (pydata["v2x_polyarchy"] < py_average_electdem)
]
print(py_filtered_data)
```

```
             country_name  year  v2x_polyarchy  e_gdppc  e_regiongeo gdp_cat
8                  Russia  2019          0.268   25.606            4    High
26            South Sudan  2019          0.174   21.725            8    High
```

```
87                  Qatar  2019          0.081   80.190           10    High
92                 Turkey  2019          0.289   23.599           10    High
100               Belarus  2019          0.257   20.806            4    High
108               Eritrea  2019          0.070   21.058            8    High
114            Kazakhstan  2019          0.237   23.827           11    High
128          Turkmenistan  2019          0.149   25.029           11    High
131               Bahrain  2019          0.118   30.046           10    High
142     Equatorial Guinea  2019          0.179   21.076            7    High
150                Kuwait  2019          0.315   45.010           10    High
154              Malaysia  2019          0.477   26.265           13    High
160                  Oman  2019          0.177   21.730           10    High
164          Saudi Arabia  2019          0.016   33.263           10    High
167             Singapore  2019          0.383   72.025           13    High
172  United Arab Emirates  2019          0.095   64.628           10    High
173               Hungary  2019          0.460   25.834            4    High
```

Q5

```r
# Find the correlation between GDP and electoral democracy rating

correlation <- cor(data$e_gdppc, data$v2x_polyarchy, use = "complete.obs", method = "pearson"
print(correlation)
```

```
[1] 0.4239001
```

```python
# Same thing in python

py_correlation = pydata["e_gdppc"].corr(pydata["v2x_polyarchy"])
print(py_correlation)
```

```
0.4239001395493961
```

Q6

While there is a moderate positive correlation between GDP per capita and electoral democracy ratings, there are many exceptions to this. There are low democracy rating countries that have high GDP per capita and they often are resource rich in oil and have authoritarian/monarchical political systems.