MATH 140   **Name** _____   **Date:**
Worksheet: US Governors 2024

The Scene

We use RStudio to investigate a data set giving information about the current governors of the 50 US States.

Begin this worksheet by completing the following steps:

1. Open RStudio.

2. Open a new script in RStudio.

3. In line 1 of your script enter and run `library(tidyverse)`

4. In line 2 of your script enter and run

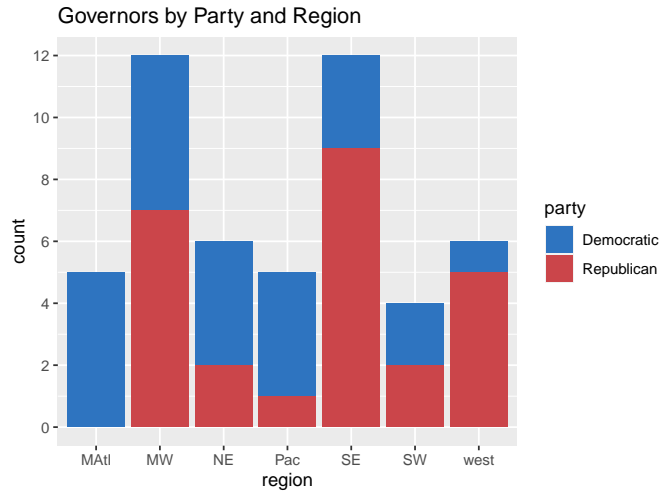   `gov = read.csv("https://mphitchman.com/stats/data/gov24.csv")`

---

1. How many observations does this data set have? How many variables? Classify each variable as numerical or categorical.

2. Use the `table()` command to determine how many current governors are Republican and how many are Democratic.

3. What proportion of current governors are Republican? Express your answer as a fraction or in decimal form.

4. Complete the two-way table of party affiliation against the `miss_riv` variable, which records whether the state is East or West of the Mississippi River. Hint: Run the `table()` command on two variables, like so: `table(gov$party,gov$miss_riv)`.

	E	W	Sum
Democratic			
Republican			
Sum			50

5. What proportion of states lie West of the Mississippi? What proportion of western states have a republican governor?

6. Below is a bar chart generated in RStudio in an effort to investigate whether there is an association between the `party` and `region` categorical variables of this data set.

   (a) How many states are in the Mid-West region (MW)? What proportion of these states has a Republican governor?

   (b) What proportion of governors in the south east (SE) is Republican?

   (c) If you pick a governor at random from the Mid-Atlantic region (MAtl), how likely is it that the governor would be Republican?

Governors by Party and Region
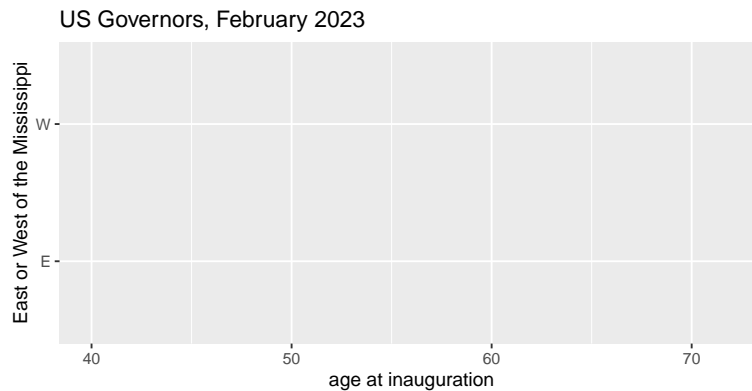
party
■ Democratic
■ Republican

7. Is there an association between a state's location and the age of its governor? Run the following code in RStudio to produce side-by-side box plots of the ages of governors East and West of the Mississippi. Sketch your answer in the blank plot below.

```
ggplot(gov,aes(x=age.at.inaug,y=miss_riv))+
  geom_boxplot(fill=c("orange","seagreen"))
```

Recall, a box plot displays a 5 number summary (Low Q1 Median Q3 High). You can compute the five number summary for the age restricted to some subset, say states East of the Mississippi, as follows:

```
fivenum(subset(gov, miss_riv=="E")$age.at.inaug)
```


US Governors, February 2023

8. Does one region *tend* to have older governors than the other? Explain in a sentence.

9. If you picked a state from the West at random, and a state from the East at random, which state would you expect to have an older governor? Explain.