Due Monday, March 26 at the start of class

Name: ______________________

Math/Stat 3850 – Take Home Quiz 2

This quiz should take you approximately 25 minutes. You may use R, the internet, and any reference material, but do not work together and do not get help (except from Dr. Clair).

The first three questions use the `babynames` dataset, which is in the `babynames` library.

1. (10) Use the `babynames` dataset to determine the most popular boys name in 1998. Then find the most popular boys name in 1999. Write your R code with your answer.

   **Solution:** In 1998, Michael. Then, in 1999, Jacob took over, knocking Michael out of the top spot for the first time since 1953.

   ```r
   babynames %>% filter(year==1998, sex=='M') %>% arrange(desc(n))
   ```

2. (10) I wish that I had Jessie’s girl. But what about Jessie? Here’s a plot showing the popularity of the name “Jessie” over time for both male and female babies:

   Reproduce this plot using the `babynames` dataset, but do it for the name “Riley”.
   You don’t need to print the plot, but describe what has happened with the name “Riley” over time - both overall and with respect to gender.

   **Solution:** Riley was a fairly uncommon boys name until about 1980, when it rapidly gained popularity, first as a boys name and then as a girls name. Around 2000, its popularity for girls surpassed its popularity for boys, and soon after that the number of male babies named Riley dropped off precipitously.

   ```r
   babynames %>% filter(name == "Riley") %>%
   ggplot(aes(x=year, y=n, color=sex)) + geom_line()
   ```
3. What are the most common letters at the start of names? Count number of babies, not number of names.

(a) Which letter is the most common first initial for boys?

(b) Find the top five most common first initials for girls.

(Hint: you can use `substr(name,1,1)` to get the first letter of a name)

**Solution:**

a. J is most common. b. In order, M, A, J, S, C.

```r
babynames %>% filter(sex=="F") %>% mutate(initial = substr(name,1,1)) %>%
group_by(initial) %>% summarize(tot=sum(n)) %>% arrange(desc(tot))
```

4. The `msleep` dataset is part of the `ggplot2` library. It has data on daily amounts of sleep needed by a selection of mammals.

(a) Which animal in this set sleeps the least?

(b) Which animal in this set sleeps the most?

(c) Which carnivore sleeps the least?

**Solution:**

5. Here is a boxplot of the `msleep` data.

Reproduce this plot with the `ggplot2` library.
Don’t print it, just write your R code here:

Solution: `ggplot(msleep, aes(x=vore, y=sleep_total)) + geom_boxplot()`