Homework 3

This homework is due on Feb. 6, 2025 at 11:00pm. Please submit as a pdf file on Canvas.

Problem 1: (8 pts) For this Problem you will be working with the `penguins2` dataset which is equivalent to `penguins` but with `NA` values removed.

```
penguins2
```

```
# A tibble: 333 × 8
   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
   <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
 1 Adelie  Torgersen           39.1          18.7               181        3750
 2 Adelie  Torgersen           39.5          17.4               186        3800
 3 Adelie  Torgersen           40.3          18                 195        3250
 4 Adelie  Torgersen           36.7          19.3               193        3450
 5 Adelie  Torgersen           39.3          20.6               190        3650
 6 Adelie  Torgersen           38.9          17.8               181        3625
 7 Adelie  Torgersen           39.2          19.6               195        4675
 8 Adelie  Torgersen           41.1          17.6               182        3200
 9 Adelie  Torgersen           38.6          21.2               191        3800
10 Adelie  Torgersen           34.6          21.1               198        4400
# i 323 more rows
# i 2 more variables: sex <fct>, year <int>
```

Use ggplot to make a histogram of the `body_mass_g` column. Manually choose appropriate values for `binwidth` and `center`. Explain your choice of values in 2-3 sentences.

```
# Your code goes here.
```

Your explanation goes here.

Problem 2: (6 pts) For Problems 2 and 3, you will work with the dataset `OH_pop` that contains Ohio state demographics and has been derived from the `midwest` dataset provided by **ggplot2**. See here for details of the original dataset: https://ggplot2.tidyverse.org/reference/midwest.html. `OH_pop` contains two columns: `county` and `poptotal` (the county's total population), and it only contains counties with at least 100,000 inhabitants.

```
OH_pop
```

```
# A tibble: 25 × 2
   county     poptotal
```

```
    <chr>          <int>
 1 CUYAHOGA     1412140
 2 FRANKLIN      961437
 3 HAMILTON      866228
 4 MONTGOMERY    573809
 5 SUMMIT        514990
 6 LUCAS         462361
 7 STARK         367585
 8 BUTLER        291479
 9 LORAIN        271126
10 MAHONING      264806
# i 15 more rows
```

Create a plot that satisfies the following two requirements:

(a) Use ggplot to make a scatter plot of `county` vs total population (column `poptotal`) and order the counties by the total population.

(b) Rename the axes and set appropriate limits, breaks and labels. Note: Do not use `xlab()` or `ylab()` to label the axes.

```
# Your code goes here.
```

Problem 2: (6 pts)

Modify the plot from Problem 2 so it satisfies the following two requirements:

(a) Change the scale for `poptotal` to logarithmic.

(b) Adjust the limits, breaks, and labels so they are appropriate for the logarithmic scale.

```
# Your code goes here.
```