431 Lab 03 Sketch and Grading Rubric

Instructor: Dr. Thomas E. Love Lab Author: Mr. Wyatt P. Bensken

Due: 2021-09-27 | Last Edit: 2021-10-02 09:10:21

Contents

1 Loading Packages

```
library(tidyverse)
```

2 Learning Objectives

1. Continue to practice and refine visualizing data in an informative way, with attention to the center, shape, and spread.
2. Use various approaches to obtain numeric summaries to assess distributions
3. Use visualization to describe the relationship between two variables.
4. Use visualizations to assess the adequacy of a Normal distribution model.

3 Packages and Functions

In Lab 03 we were hoping you would become more familiar with the following packages and functions:

Packages:

- `tidyverse`
- `palmerpenguins`
- `skimr` (optional)
- `mosaic` (optional)
- `psych` (optional)
- `ggridges` (optional)

Functions:

- `%>%`
- `ggplot()`
- `slice_sample()`
- `skim()` (optional)
- `favstats()` (optional)
- `describe()` (optional)

4 The Data for Lab 03

There are two sets of data used in this lab.

4.0.1 The `penguins` data

First, we'll be using the `penguins` data (note: use the `penguins` tibble, and not the `penguins_raw` tibble for this Lab) contained in the `palmerpenguins` package in R. The complete citation is …

Horst AM, Hill AP, Gorman KB (2020). palmerpenguins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0. https://allisonhorst.github.io/palmerpenguins/. doi: 10.5281/zenodo.3960218.

Additional information on the data are provided by Allison Horst at the github site linked above. In particular, you'll find a nice cartoon of the three species of penguin contained in the data and a detailed description of the bill measurements that are worth your time.

You'll first want to load in the penguins data, as shown below.

```r
penguins <- palmerpenguins::penguins
```

4.0.2 County Health Rankings data

Next, we'll be using additional data from the 2021 County Health Rankings Data that we saw back in Lab 02. We have compiled a larger data set called `lab03_counties.csv` which you will find in our Data and Code repository. This data file contains all of the counties and a dozen of the available variables in the full County Health Rankings data.

We will load in our CSV file.

```r
lab03_chr <- read_csv("data/lab03_counties.csv")
```

```
Rows: 3142 Columns: 12
```

```
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (6): state, county_name, teen_births, flu_vaccinations, life_expectancy,...
dbl (6): fips, adult_smoking, adult_obesity, some_college, food_insecurity, ...
```

```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

5 Part A: Palmer Penguins (40 points total, 8 points each)

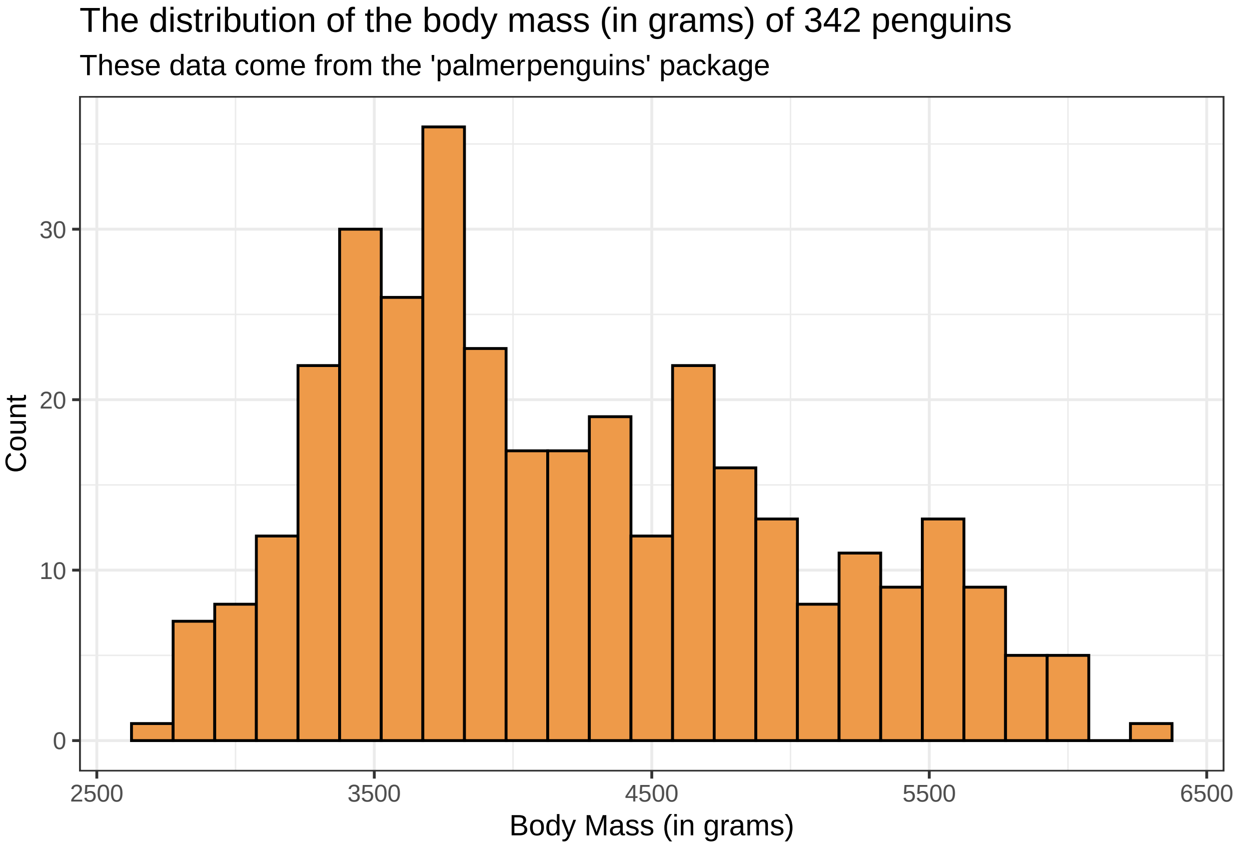
5.1 Question 1

> We are interested in the body mass (in grams), `body_mass_g`, of the penguins included in these data. Create a visualization of this variable that will help us evaluate its center, shape, and spread, using appropriate labels for all axes, and a useful title for the visualization.

The best visualization for this would likely be a histogram. As with any visualization we build in this class, we should modify the defaults to make it more interpretable and useful. Additionally, we should get a warning from these visualizations that two rows were removed - so we only have 342 included. We can identify those by filtering to those rows with missing values (see Comments).

```r
ggplot(data = penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 150, colour = "black", fill = "tan2") +
  labs(x = "Body Mass (in grams)",
       y = "Count",
       title = "The distribution of the body mass (in grams) of 342 penguins",
       subtitle = "These data come from the 'palmerpenguins' package") +
  theme_bw()
```

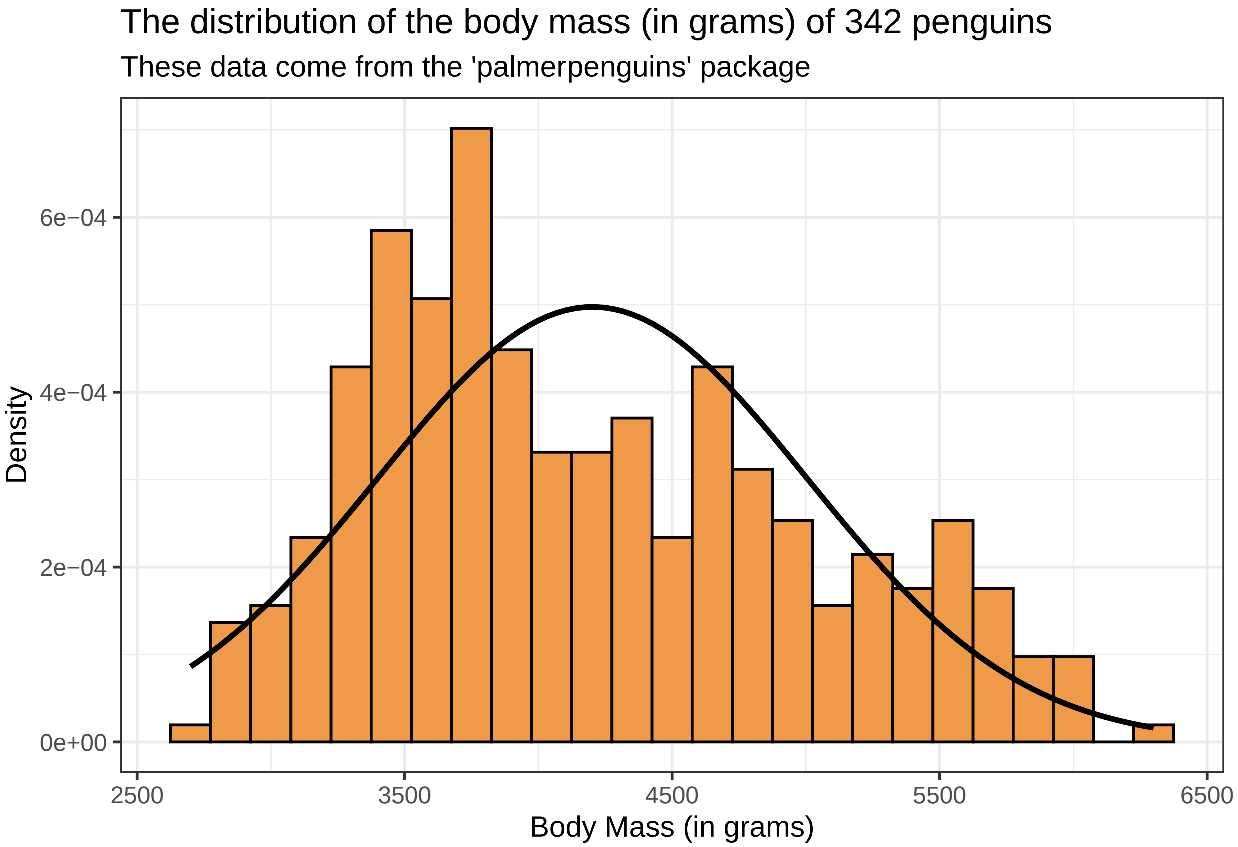
```
Warning: Removed 2 rows containing non-finite values (stat_bin).
```



Anticipating question 3, we can also build a visualization which puts a Normal distribution over our histogram. An important note here is we change our y-axis from a count to a density.

```r
ggplot(data = penguins, aes(x = body_mass_g)) +
  geom_histogram(aes(y = ..density..),
                 binwidth = 150,
                 colour = "black",
                 fill = "tan2") +
  stat_function(fun = dnorm,
                args = list(mean = mean(penguins$body_mass_g, na.rm = TRUE),
                            sd = sd(penguins$body_mass_g, na.rm = TRUE)),
                col = "black",
                size = 1) +
  labs(x = "Body Mass (in grams)",
       y = "Density",
       title = "The distribution of the body mass (in grams) of 342 penguins",
       subtitle = "These data come from the 'palmerpenguins' package") +
  theme_bw()
```

```
Warning: Removed 2 rows containing non-finite values (stat_bin).
```

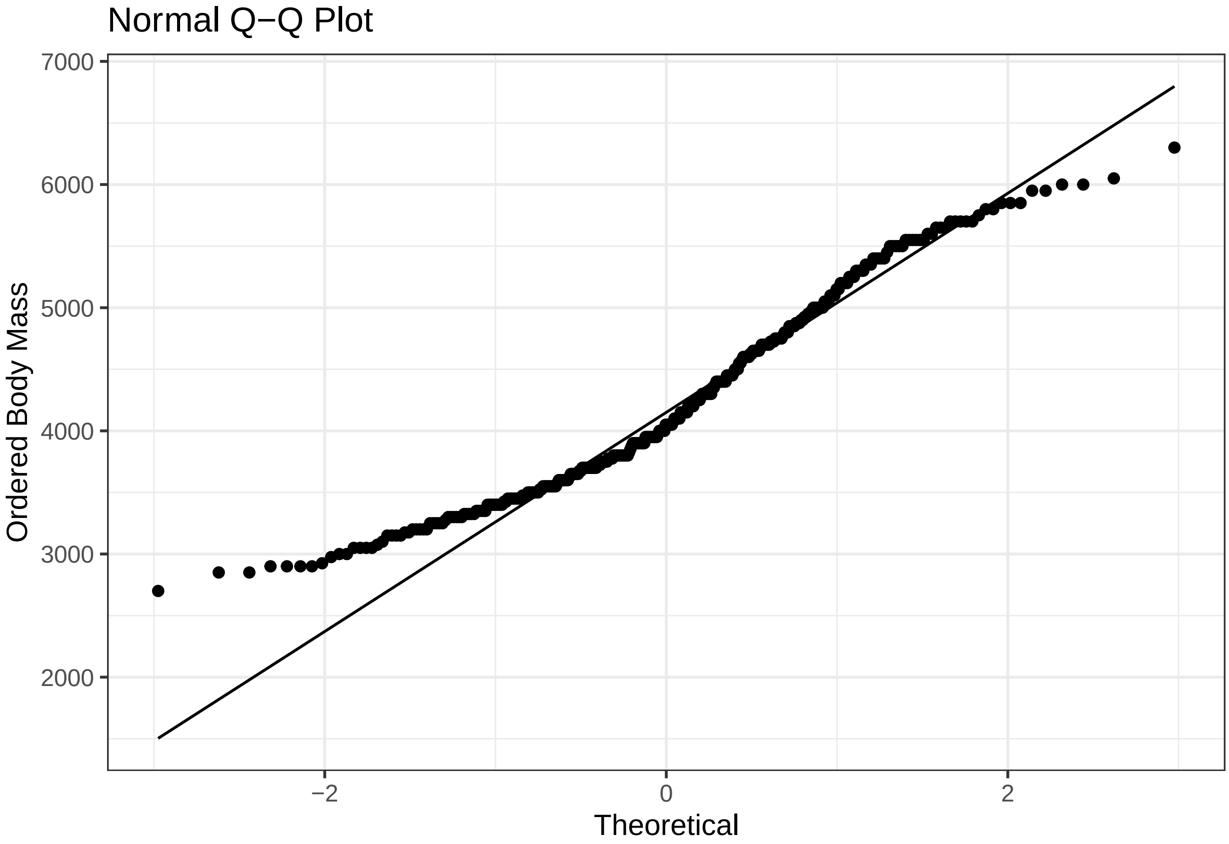


Another important approach to consider is using a Normal Q-Q plot to assess how well these data fit a Normal distribution. Our conclusion here should not change, but rather it becomes a bit easier to see the skew that we have in these data.

```
ggplot(data = penguins, aes(sample = body_mass_g)) +
  geom_point(stat = "qq") +
  geom_qq_line() +
  labs(x = "Theoretical",
       y = "Ordered Body Mass",
       title = "Normal Q-Q Plot") +
  theme_bw()
```

```
Warning: Removed 2 rows containing non-finite values (stat_qq).
```

```
Warning: Removed 2 rows containing non-finite values (stat_qq_line).
```

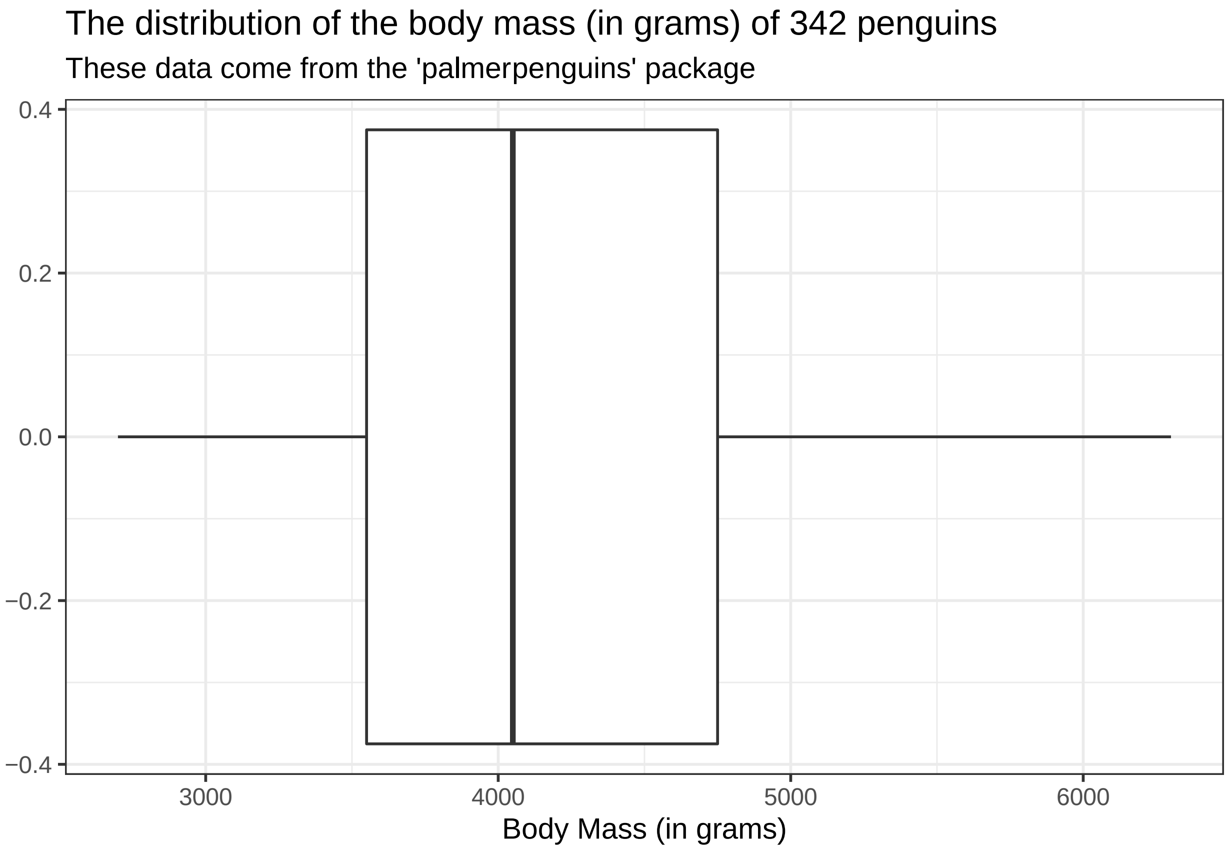


5.1.1 Comments

One potential alternative, although less useful, would be a boxplot.

```
ggplot(data = penguins, aes(x = body_mass_g)) +
  geom_boxplot() +
  labs(x = "Body Mass (in grams)",
       title = "The distribution of the body mass (in grams) of 342 penguins",
       subtitle = "These data come from the 'palmerpenguins' package") +
  theme_bw()
```

```
Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```



5.1.1.1 Missing Data As noted, we have 2 rows with missing data. We can see them by filtering to them. For simplicity sake, and to prevent the warning from continuing, we will go ahead and remove them.

```
penguins %>%
  filter(is.na(body_mass_g))
```

```
# A tibble: 2 x 8
  species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g sex
  <fct>   <fct>           <dbl>         <dbl>            <int>       <int> <fct>
1 Adelie  Torge~             NA            NA               NA          NA <NA>
2 Gentoo  Biscoe             NA            NA               NA          NA <NA>
# ... with 1 more variable: year <int>
```

```
penguins <- penguins %>%
  filter(!is.na(body_mass_g))
```

5.1.2 Grading Rubric

- If the student has produced a visually pleasing figure, with a reasonable title and axes labels, which are not just the variable names, they should receive all 8 points.
- If any of the following occur the student should lose 2 points for each:
  - No title, or a title which does not convey what the figure represents
  - Axes remain unlabeled or with the default labels
- If the figure provided does not allow one to interpret the center, shape, and spread the student should lose up to 5 points.

5.2 Question 2

> Identify the mean, standard deviation, median and interquartile range of the penguin's body mass (in grams). Do not simply report R output but (in addition to presenting your code and the resulting output) state, in complete English sentences, your results.

There are a number of ways to identify these standard numeric summary variables. A few of the simplest include using the `skim()` function from the `skimr` package. From these results we can see:

- a mean of 4,202
- a standard deviation of 802
- a median (P50) of 4,050
- an interquartile range of 3,550 to 4,750

```
penguins %>%
  select(body_mass_g) %>%
  skimr::skim_without_charts()
```

Table 1: Data summary

Name	Piped data
Number of rows	342
Number of columns	1
Column type frequency:	
numeric	1
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
body_mass_g	0	1	4201.75	801.95	2700	3550	4050	4750	6300

5.2.1 Comments

There are numerous alternatives for this summary, including:

Using `favstats()` from the `mosaic` package.

```
mosaic::favstats(penguins$body_mass_g)
```

```
Registered S3 method overwritten by 'mosaic':
  method                           from
  fortify.SpatialPolygonsDataFrame ggplot2
```

```
  min   Q1 median   Q3  max     mean       sd   n missing
 2700 3550   4050 4750 6300 4201.754 801.9545 342       0
```

There are also two versions of `describe()`, one from the `psych` package and one from the `hmisc` package.

```
psych::describe(penguins$body_mass_g)
```

```
   vars   n    mean     sd median trimmed    mad  min  max range skew kurtosis
```

```
X1    1 342 4201.75 801.95   4050 4154.01 889.56 2700 6300  3600 0.47    -0.74
      se
X1 43.36
```

```
Hmisc::describe(penguins$body_mass_g)
```

```
penguins$body_mass_g
       n  missing distinct     Info     Mean      Gmd      .05      .10
     342        0       94        1     4202    911.8     3150     3300
     .25      .50      .75      .90      .95
    3550     4050     4750     5400     5650

lowest : 2700 2850 2900 2925 2975, highest: 5850 5950 6000 6050 6300
```

5.2.2 Grading Rubric

- If the student has correctly identified the mean, standard deviation, median, and interquartile range and reported these values in complete English sentences, they should receive all 8 points.
- If the values are correct, but reported simple as R output (and not sentence format) the student should lose 4 points.
- If the values are incorrect, but reported in sentence format the student should lose 4 points.
- If the values are incorrect and are not reported in sentence format the student should lose 8 points.

5.3 Question 3

> Given your visualization in Question 1 and the numeric summary in Question 2, please discuss the center, shape, and spread of the body mass of the penguins. Does it seem that body mass follows a Normal distribution? Would it be more appropriate to examine the mean or median in this setting, and why?

Overall, we see the center of our distribution right around 4,000 grams. When we overlay a Normal distribution we see its shape does not exactly follow a Normal distribution, but is relatively close. As for the spread, we can see in the visualization some light right-skew to these data (meaning a higher number of lower body mass penguins), which is confirmed with our mean lower than our median. We can also look at the standard deviation of 801.95 and the interquartile range of [3550, 4750] as other measures of our spread. Given all of this, we may be more interested in the median than the mean as a measure of central tendency. From the numeric summary, we see that the mean and median are somewhat close. We could certainly model the data using a model which assumes a Normal distribution, but it will be important to pay attention to checking the model assumptions.

5.3.1 Grading Rubric

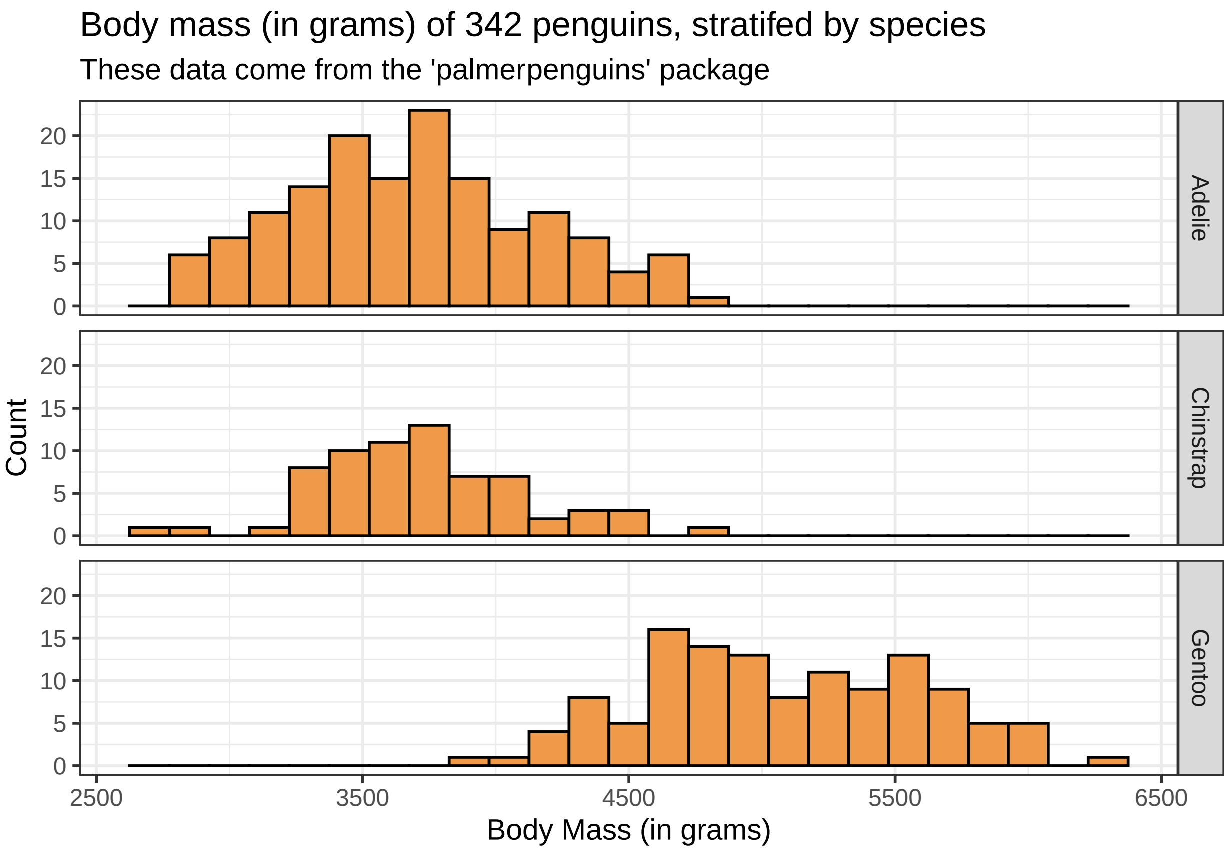
- If a student a) discusses the center, shape, and spread, b) identifies a moderate amount of skew, and c) suggests the median may be more appropriate they should receive they should receive the full 8 points.
  - For each of these not covered, the student should lose 2 points.

5.4 Question 4

> The dataset also contains information on the penguin's species (`species`). Now, build a visualization and a numeric summary to examine body mass across the three species types.

There are many different options to accomplish this visualization. The first, and likely simplest, is to just facet our first histogram by species.

```
ggplot(data = penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 150, colour = "black", fill = "tan2") +
  facet_grid(species ~ .) +
  labs(x = "Body Mass (in grams)",
       y = "Count",
       title = "Body mass (in grams) of 342 penguins, stratifed by species",
       subtitle = "These data come from the 'palmerpenguins' package") +
  theme_bw()
```



We can expand our numeric summary to get the means, standard deviations, medians, and interquartile range stratified by species.

```
penguins %>%
  select(body_mass_g, species) %>%
  group_by(species) %>%
  skimr::skim_without_charts()
```

Table 3: Data summary

Name	Piped data

Number of rows	342
Number of columns	2
Column type frequency:	
numeric	1
Group variables	species

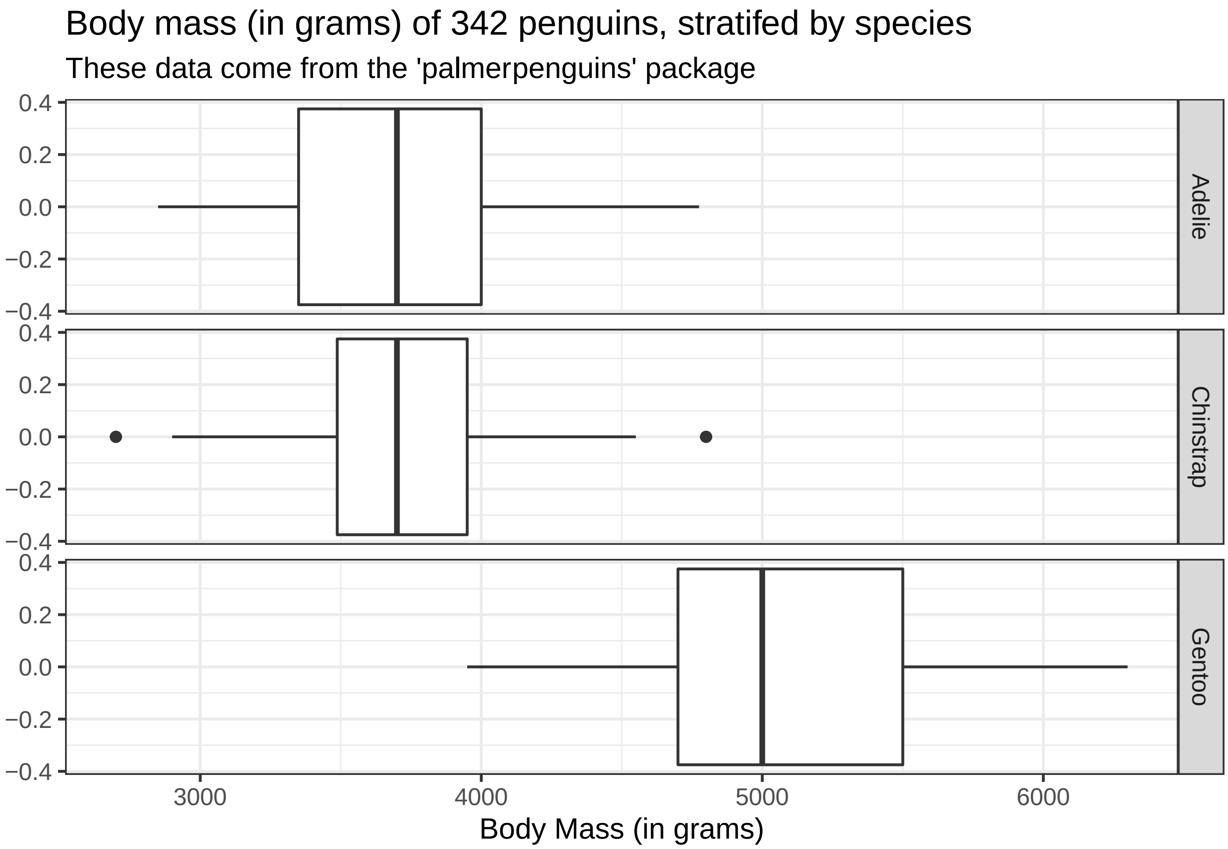
Variable type: numeric

skim_variable	species	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
body_mass_g	Adelie	0	1	3700.66	458.57	2850	3350.0	3700	4000	4775
body_mass_g	Chinstrap	0	1	3733.09	384.34	2700	3487.5	3700	3950	4800
body_mass_g	Gentoo	0	1	5076.02	504.12	3950	4700.0	5000	5500	6300

5.4.1 Comments

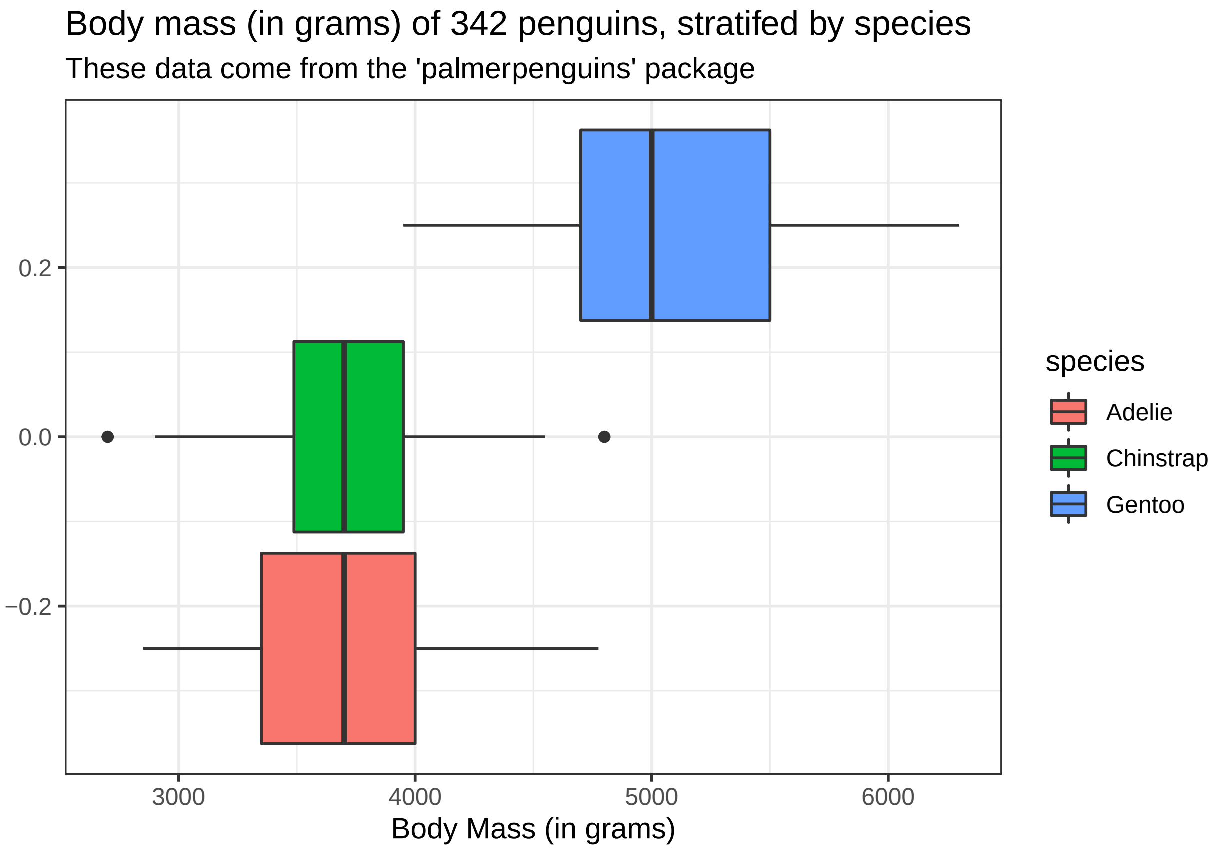
Similarly, we can also facet our boxplot.

```r
ggplot(data = penguins, aes(x = body_mass_g)) +
  geom_boxplot() +
  facet_grid(species ~ .) +
  labs(x = "Body Mass (in grams)",
       title = "Body mass (in grams) of 342 penguins, stratifed by species",
       subtitle = "These data come from the 'palmerpenguins' package") +
  theme_bw()
```



Alternatively, we could instead add species as a fill which adds some colors and slightly changes how the visualization appears.

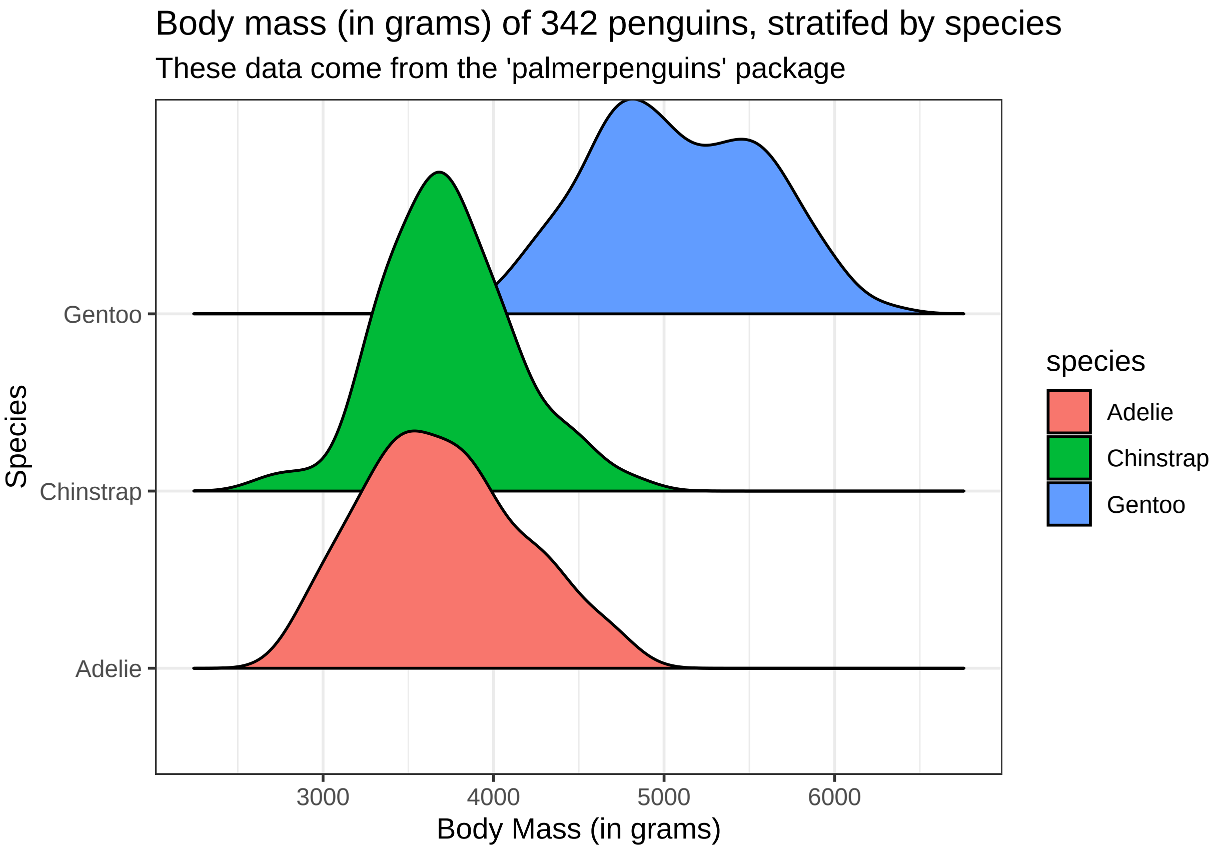
```
ggplot(data = penguins, aes(x = body_mass_g, group = species)) +
  geom_boxplot(aes(fill = species)) +
  labs(x = "Body Mass (in grams)",
       title = "Body mass (in grams) of 342 penguins, stratifed by species",
       subtitle = "These data come from the 'palmerpenguins' package") +
  theme_bw()
```



Finally, we could build upon our histogram by using what's called a ridgeline plot from the `ggridges` package.

```
ggplot(data = penguins, aes(x = body_mass_g, y = species, fill = species)) +
  ggridges::geom_density_ridges2() +
  labs(x = "Body Mass (in grams)",
       y = "Species",
       title = "Body mass (in grams) of 342 penguins, stratifed by species",
       subtitle = "These data come from the 'palmerpenguins' package") +
  theme_bw()
```

```
Picking joint bandwidth of 153
```



We can also use `mosaic` and `psych` to get the stratified numeric summaries.

```
mosaic::favstats(body_mass_g ~ species, data = penguins)
```

```
    species  min     Q1 median   Q3  max     mean       sd   n missing
1    Adelie 2850 3350.0   3700 4000 4775 3700.662 458.5661 151       0
2 Chinstrap 2700 3487.5   3700 3950 4800 3733.088 384.3351  68       0
3    Gentoo 3950 4700.0   5000 5500 6300 5076.016 504.1162 123       0
```

```
psych::describeBy(penguins$body_mass_g, group = penguins$species, mat = TRUE)
```

```
    item    group1 vars   n     mean       sd median  trimmed     mad  min  max
X11    1    Adelie    1 151 3700.662 458.5661   3700 3685.744 444.780 2850 4775
X12    2 Chinstrap    1  68 3733.088 384.3351   3700 3719.643 370.650 2700 4800
X13    3    Gentoo    1 123 5076.016 504.1162   5000 5073.485 555.975 3950 6300
    range       skew   kurtosis       se
X11  1925 0.27969223 -0.6261455 37.31758
X12  2100 0.23662398  0.3625567 46.60747
X13  2350 0.06794565 -0.7786932 45.45463
```

5.4.2 Grading Rubric

- If the student has produced a visually pleasing figure, with a reasonable title and axes labels, which can answer the question posed they should receive all 8 points.
- If any of the following occur the student should lose 2 points for each:
  - No title, or a title which does not convey what the figure represents
  - Axes remain unlabeled or with the default labels

- If the figure provided does not allow one to compare the distribution by species, the student should lose up to 6 points.
- If the student does not also display a numeric summary, they should lose 2 points.

5.5 Question 5

Given your findings in Question 4, what can we conclude about the body mass across species?

Regardless of our visualization choice or our numeric summary, however, we can clearly see there is meaningful variation in the body mass of the penguins across the 3 species. Specifically, it seems the Gentoo penguins have a higher body mass, with the Adelie and Chinstrap quite close to each other. Our numeric summary reveals that the mean is slightly higher the Chinstrap penguins than the Adelie. However, the two species have the same median. Overall, it seems as though the species of the penguin is important when considering its body mass.

5.5.1 Grading Rubric

- If a student concludes that the weight of the penguins varies by species, and specifies the direction of this variation (i.e. which species is highest and lowest) they should receive the full 8 points.
  - If a student does not specify the direction of variations they should lose 2 points.

6 Part B: County Health Rankings (40 points total, 10 points each)

6.1 Question 6

Take a random sample of 750 counties from the County Health Rankings Data provided in `lab03_counties.csv`. As part of this work, use the command:

```
set.seed(20212022)
```

so that each of us selects the same sample of data. Name this random sample `chr_sample` in R. Select only the following variables: `state`, `county_name`, `adult_obesity`, and `food_insecurity`.

Once this is done, demonstrate that Cuyahoga County in Ohio is in your sample and that the mean of `adult_obesity` is 0.3345.

```
set.seed(20212022)
chr_sample <- slice_sample(lab03_chr, n = 750)

chr_sample <- chr_sample %>%
  select(county_name, adult_obesity, food_insecurity)
```

Now we want to verify 2 things. First that Cuyahoga County does, in fact, appear in our sample.

```
chr_sample %>%
  filter(county_name == "Cuyahoga County")
```

```
# A tibble: 1 x 3
  county_name     adult_obesity food_insecurity
  <chr>                   <dbl>           <dbl>
1 Cuyahoga County         0.318           0.159
```

Second, that the mean of the proportion of adults who are obese is 0.3345027.

```
mean(chr_sample$adult_obesity)
```

```
[1] 0.3345027
```

6.1.1 Grading Rubric

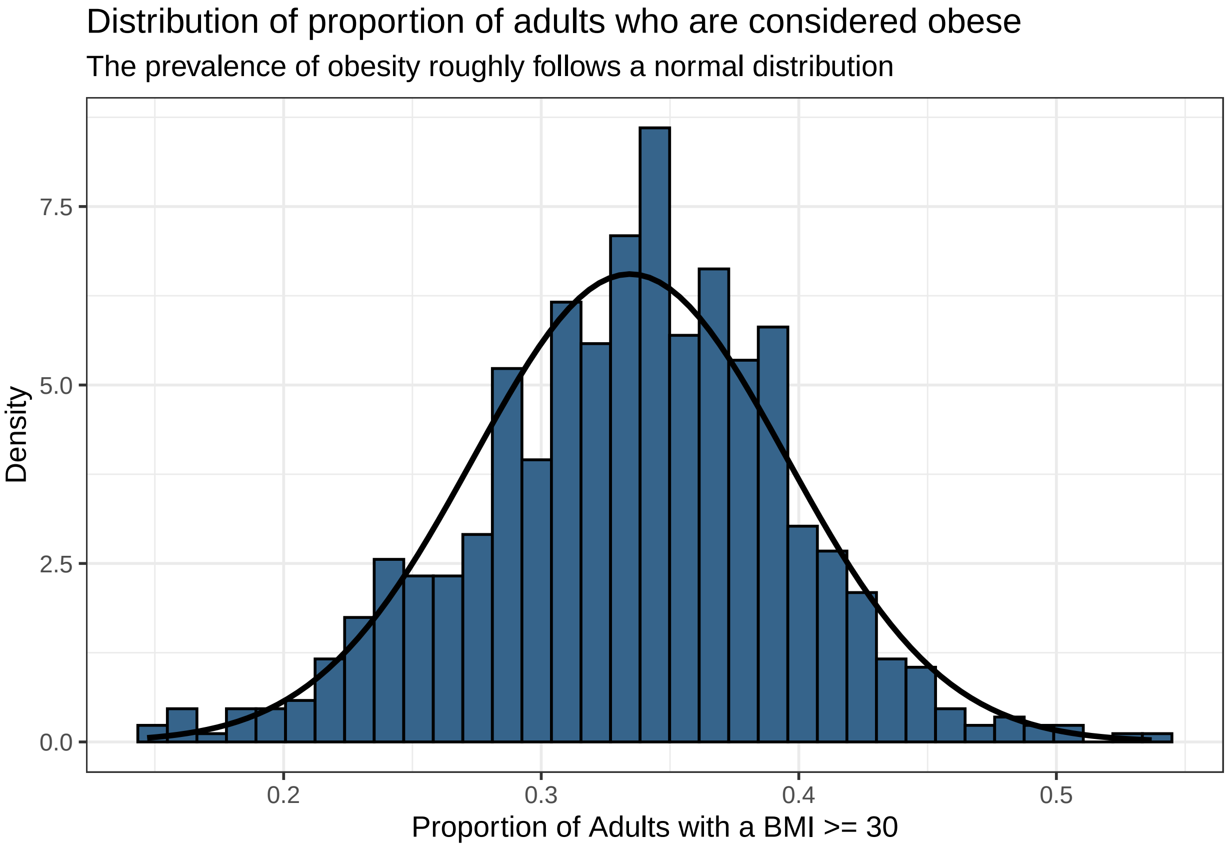
- If a student successfully took a subset of the data, has Cuyahoga county in the sample, and the correct mean they should receive the full 10 points.
  - If the student did not use R code to confirm that Cuyahoga County was in the sample and/or the mean of `adult_obesity` they should lose 4 points.

6.2 Question 7

> We are interested in looking at `adult_obesity` as an outcome. Build a visualization and describe the center, spread, and shape of the distribution of this variable. Does a Normal model for this distribution seem appropriate? Why or why not?

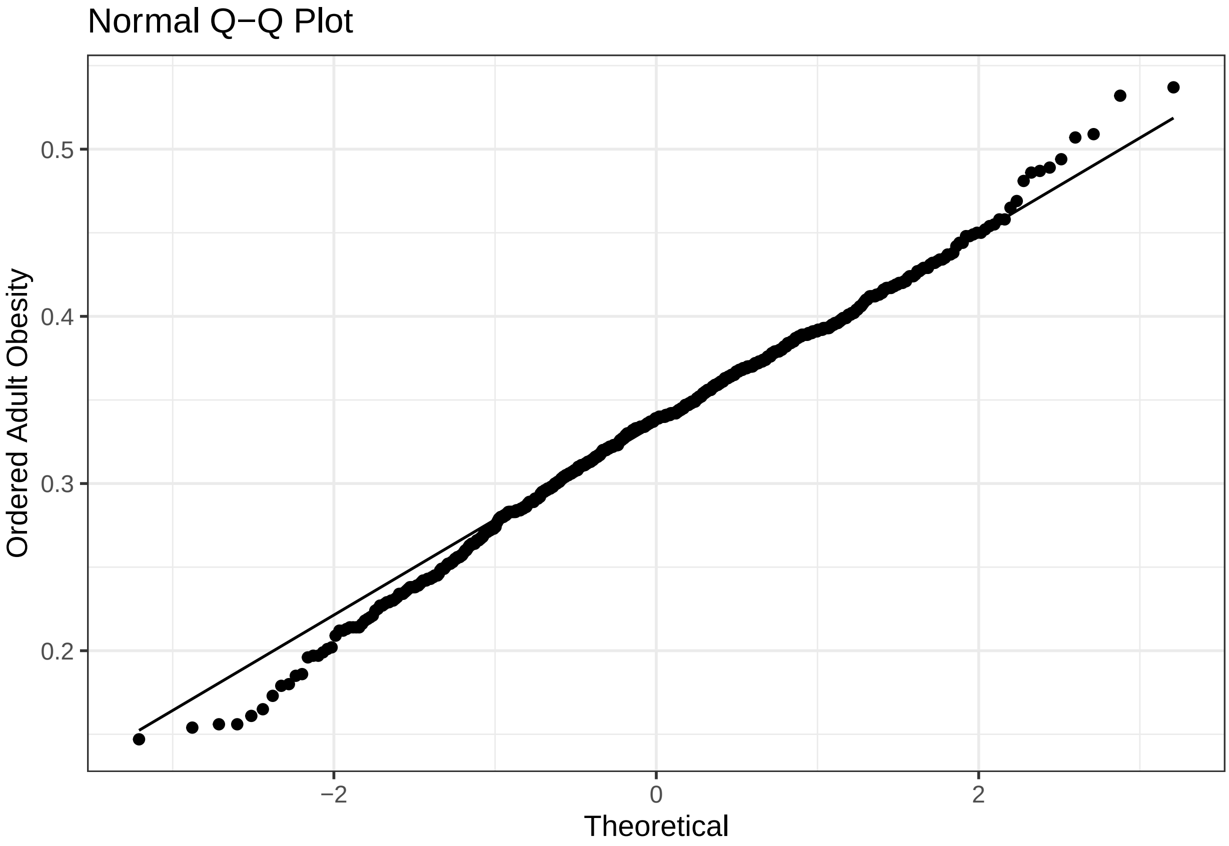
In the visualization below we have built a histogram, with the Normal distribution overlaid. We can see from this figure that our outcome of interest here, follows a Normal distribution quite closely. We observe a center right around 4.5 days (per month).

```
ggplot(data = chr_sample, aes(x = adult_obesity)) +
  geom_histogram(aes(y = ..density..),
                 fill = "steelblue4",
                 colour = "black",
                 bins = 35) +
  stat_function(fun = dnorm,
                args = list(mean = mean(chr_sample$adult_obesity, na.rm = TRUE),
                            sd = sd(chr_sample$adult_obesity, na.rm = TRUE)),
                col = "black",
                size = 1) +
  labs(x = "Proportion of Adults with a BMI >= 30",
       y = "Density",
       title = "Distribution of proportion of adults who are considered obese",
       subtitle = "The prevalence of obesity roughly follows a normal distribution") +
  theme_bw()
```



As we have previously discussed, a Normal Q-Q plot could also help us visualize and determine the appropriateness of a Normal distribution model.

```r
ggplot(data = chr_sample, aes(sample = adult_obesity)) +
  geom_point(stat = "qq") +
  geom_qq_line() +
  labs(x = "Theoretical",
       y = "Ordered Adult Obesity",
       title = "Normal Q-Q Plot") +
  theme_bw()
```



6.2.1 Grading Rubric

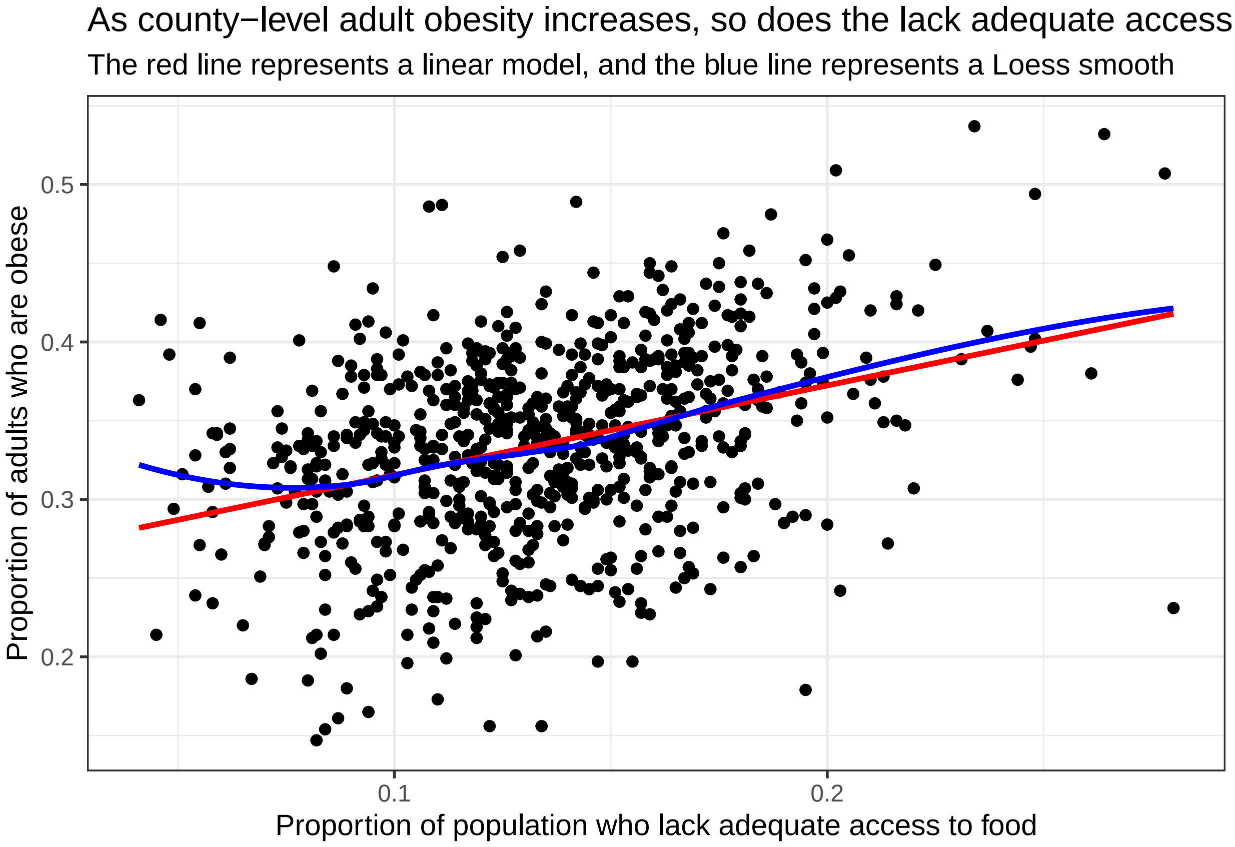
- If the student has produced a visually pleasing figure, with a reasonable title and axes labels, which are not just the variable names, they should receive all 10 points.
- If any of the following occur the student should lose 2 points for each:
    - No title, or a title which does not convey what the figure represents
    - Axes remain unlabeled or with the default labels
- Students will lose 1 point if the axis label is another version of the variable name and not descriptive.
- If the figure provided does not allow one to interpret the center, shape, and spread the student should lose up to 5 points.
- If the student incorrectly concludes that this does not follow a Normal distribution, they should lose 3 points.

6.3 Question 8

> Now we’d like to examine the relationship between `food_insecurity` (our predictor) and `adult_obesity` (our outcome). Build a useful, well-labeled visualization and then use it to motivate your description of the relationship between these two variables in complete sentences. Does a linear model seem appropriate to describe the association of these variables? Why or why not? Hint: you will likely want to include a smooth of some sort in your visualization.

In the visualization below, we see that generally as we move from left to right on the x-axis (i.e., the proportion of the population who lack adequate access to food increases), we also tend to move from down to up on the y-axis (i.e., the proportion of adults who are obese also increases). It seems there is a moderate positive relationship. This is confirmed with our our linear model line (in red) and a Loess smooth line (in blue). Further, while the Loess does have some bend at lower values of `food_insecurity`, overall it’s quite similar to the straight line. What is notable, too, is that the spread of these points around the straight line suggest that while there is a positive relationship, it may not be all the strong. Overall, though, it does seem as though a linear model is appropriate.

```
ggplot(data = chr_sample, aes(x = food_insecurity, y = adult_obesity)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, formula = y ~ x, col = "red") +
  geom_smooth(method = "loess", se = FALSE, formula = y ~ x, col = "blue") +
  labs(x = "Proportion of population who lack adequate access to food",
       y = "Proportion of adults who are obese",
       title = "As county-level adult obesity increases, so does the lack adequate access to food",
       subtitle = "The red line represents a linear model, and the blue line represents a Loess smooth")
  theme_bw()
```



6.3.1 Grading Rubric

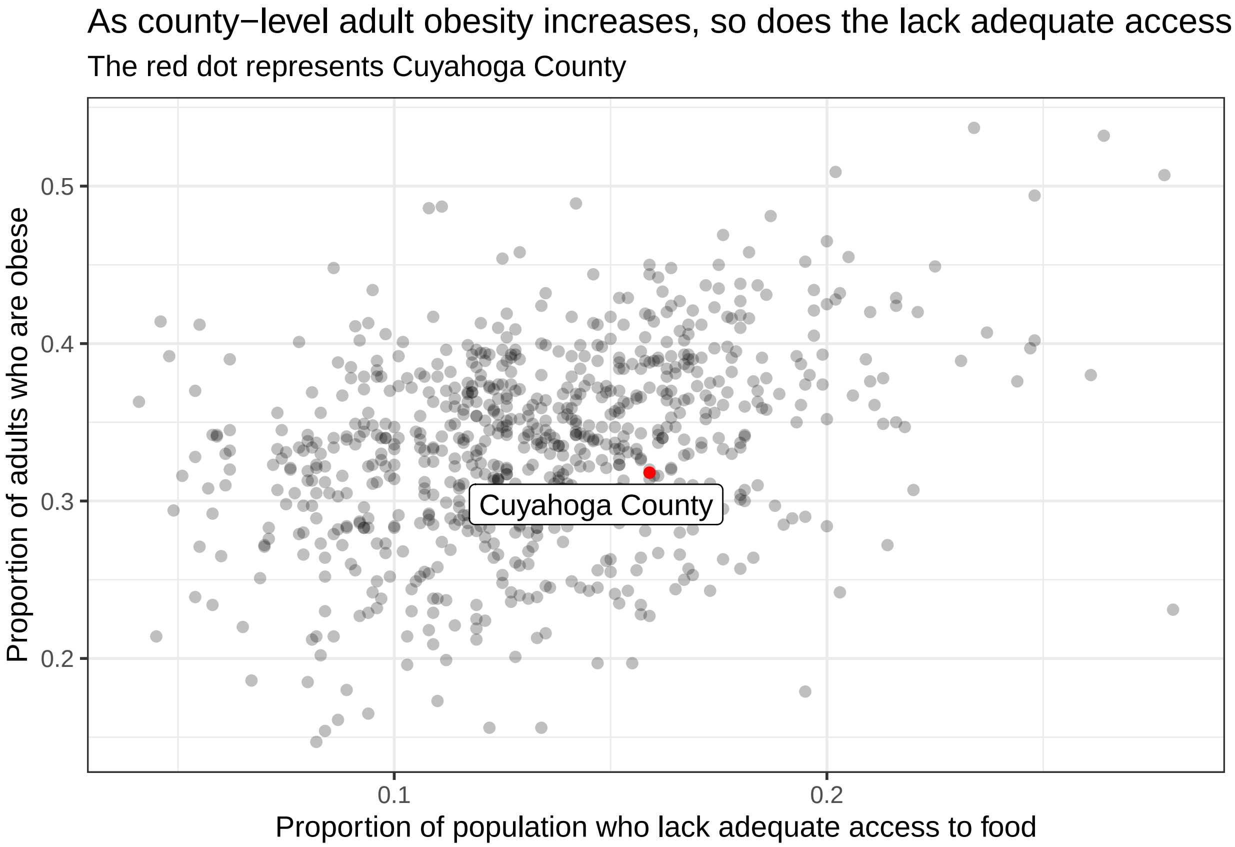
- If a student has an appropriate visualization, with a linear model line as well as a Loess smooth curve, and accurately concludes that a linear model is appropriate they should receive all 10 points.
  - If the student states that a linear model is appropriate but has not applied a linear model to their figure they should lose 3 points.
  - If any of the following occur the student should lose 2 points for each:
    - No title, or a title which does not convey what the figure represents
    - Axes remain unlabeled or with the default labels
- If the student does not interpret the relationship, but has a figure which allows one to, they should lose 4 points.
- If a student concludes a linear model is not appropriate, they should lose 3 points.
  - An additional 3 points should be deducted if they do not appropriately justify this decision.

6.4 Question 9

> Create a new figure (perhaps building on the one you built in Question 8) so that this new Figure identifies where Cuyahoga County falls when examining the relationship between `adult_obesity` and `food_insecurity`. Briefly interpret Cuyahoga County's position on each variable relative to the others in the random sample.

In the figure below, we have done three things. The first is to make all of the other points (Counties) transparent, the second is to color Cuyahoga County red, and the third is to add a label. As we can see from this point Cuyahoga County falls a bit to the right and lower than the middle. This would suggest a higher proportion of adults who lack adequate access to food and lower prevalence of obesity.

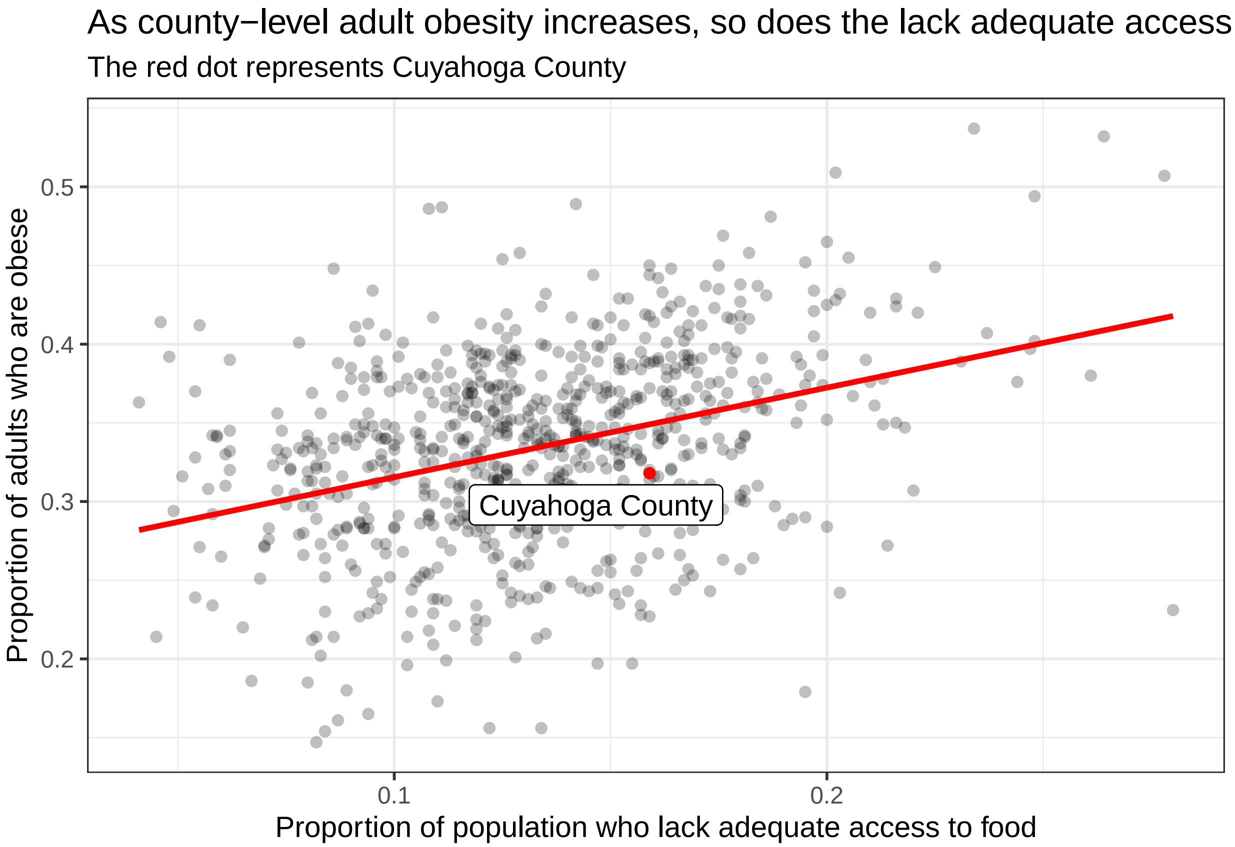
```r
ggplot(data = chr_sample, aes(x = food_insecurity, y = adult_obesity)) +
  geom_point(alpha = 0.25) +
  geom_point(data = chr_sample %>% filter(county_name == "Cuyahoga County"), colour = "red") +
  ggrepel::geom_label_repel(
    data = subset(chr_sample,
                  county_name == "Cuyahoga County"),
    mapping = aes(label = "Cuyahoga County"),
    min.segment.length = unit(100, 'lines')) +
  labs(x = "Proportion of population who lack adequate access to food",
      y = "Proportion of adults who are obese",
      title = "As county-level adult obesity increases, so does the lack adequate access to food",
      subtitle = "The red dot represents Cuyahoga County") +
  theme_bw()
```



6.4.1 Comments

We can further our understanding of Cuyahoga County's position by adding the linear model fit as we did in Question 4. By adding this line we can see that if we were to use this linear model to "predict" Cuyahoga County's obesity prevalence from the proportion of adults who lack adequate access to food, we would overestimate the prevalence of obesity.

```r
ggplot(data = chr_sample, aes(x = food_insecurity, y = adult_obesity)) +
  geom_point(alpha = 0.25) +
  geom_point(data = chr_sample %>% filter(county_name == "Cuyahoga County"),
             colour = "red") +
  geom_smooth(method = "lm", se = FALSE, formula = y ~ x, col = "red") +
  ggrepel::geom_label_repel(
    data = subset(chr_sample,
                  county_name == "Cuyahoga County"),
    mapping = aes(label = "Cuyahoga County"),
    min.segment.length = unit(100, 'lines')) +
  labs(x = "Proportion of population who lack adequate access to food",
       y = "Proportion of adults who are obese",
       title = "As county-level adult obesity increases, so does the lack adequate access to food",
       subtitle = "The red dot represents Cuyahoga County") +
  theme_bw()
```



6.4.2 Grading Rubric

- If a student has successfully identified Cuyahoga County, either with a label or a different colored point, or another mechanism, on the figure, they should receive all 10 points.

- If the student's figure does not show the relationship between `adult_obesity` and `food insecurity`, but instead they identify Cuyahoga county's values on these variables separately, they should lose 4 points.
- If any of the following occur the student should lose 2 points for each:
  - No title, or a title which does not convey what the figure represents
  - Axes remain unlabeled or with the default labels

7 Part C: Spiegelhalter Reaction (20 points)

7.1 Question 10

> Reflecting on Chapter 4 of *The Art of Statistics*, please write an essay of no more than 100 words which discusses the relationship between `adult_obesity` and `food_insecurity` we observe in our sample of counties. Specifically, discuss whether or not we can conclude that lack of access to adequate food causes obesity.
>
> In your response be sure to also discuss whether the method that we used to identify these data (our 750 observation random sample) strengthens or weakens your conclusion(s).

We don't write sketches for essay questions.

7.1.1 Grading Rubric

For essay questions, each student starts with 20 full points, and deductions should be noted in the TA notes of the grading sheet.

A 20-point answer will include a discussion of how these analyses alone are not enough to draw any strong conclusions regarding the relationship between food insecurity and obesity. While these exploratory analyses are helpful, there is certainly not enough to draw any causal relationship. Further, it would be helpful to note that this analysis may be more generalizable and more compelling given that we took a random sample from all of our counties, rather than selecting specific counties or states, as we have done previously. To receive full credit you must directly tie these observations to Spiegelhalter.

- If a student covers most of these topics then a score of 20 points is appropriate.
- If a student discusses a smaller number of these topics, but in greater depth, than a score of 19 or 20 is appropriate.
- If a student fails to discuss these topics but reasonably relates to concepts covered in Spiegelhalter, they should score 15 to 17.
- If a student states that these analyses are sufficient to draw conclusions they should receive a score of 12 - 15.
  - If the student uses causal language they should receive a score of 10 - 1
- If a student does not relate their response to Spiegelhalter, does not discuss any of these concepts, but does have an essay, they should score 8 to 10.
- If no essay is provided a student should receive 0 points and "No Essay" should be noted on the grading sheet.