Homework 1

```
library(tidyverse)
theme_set(theme_bw())
```

Question 1 [30 Points total]

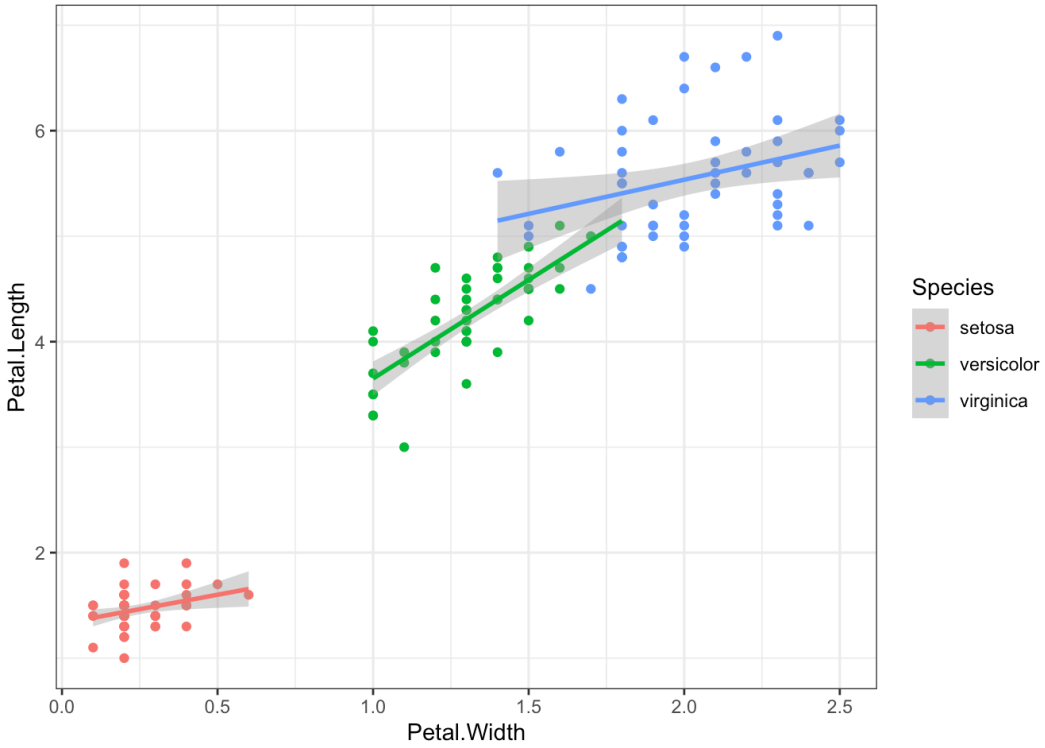
Question 1(a) [10 Points]

The iris dataset is available in R by default.

```
iris |>
  glimpse()
```

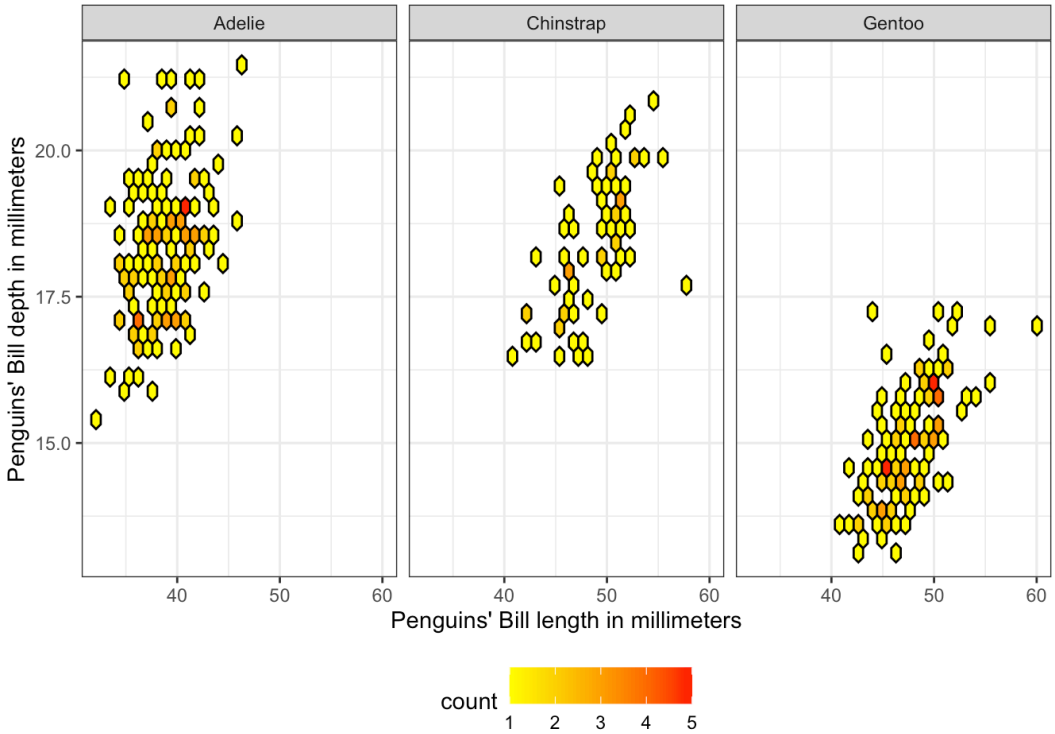
```
Rows: 150
Columns: 5
$ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0,
4.4, 4.9, 5.4, 4.…
$ Sepal.Width  <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4,
2.9, 3.1, 3.7, 3.…
$ Petal.Length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5,
1.4, 1.5, 1.5, 1.…
$ Petal.Width  <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2,
0.2, 0.1, 0.2, 0.…
$ Species      <fct> setosa, setosa, setosa, setosa, setosa,
setosa, setosa, s…
```

Please use this dataset to create the following plot.



Question 1(b) [10 Points]

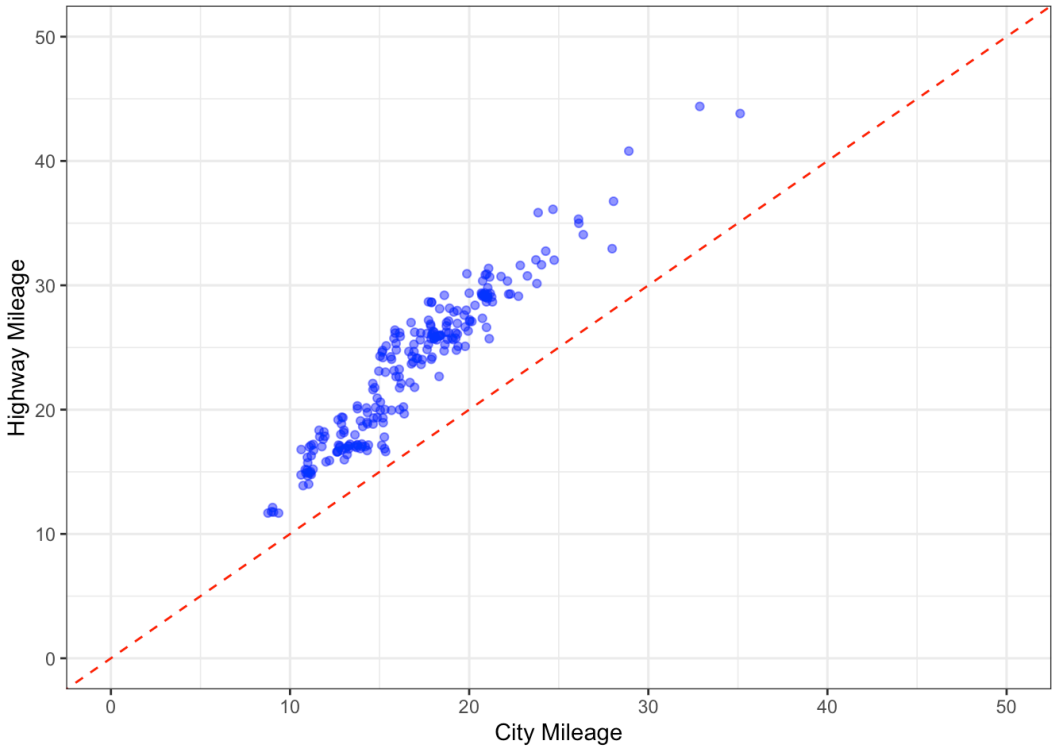
Use ggplot2 and the penguins dataset to recreate the following visualization.



Question 1(c) [10 Points]

Use ggplot2 and the mpg dataset to approximately recreate the following scatterplot. It demonstrates that all cars in the mpg dataset get better highway mileage than city mileage.

Hint: The alpha level (transparency) used is 0.5, and the points are jittered using a seed of 1. Info on creating the dotted line with intercept 0 and slope 1 is available in the extras.qmd document in HuskyCT.



Question 2 [50 Points Total]

This question will apply to the gapminder dataset, available within the gapminder package on CRAN. Please download and load this package to complete the following questions.

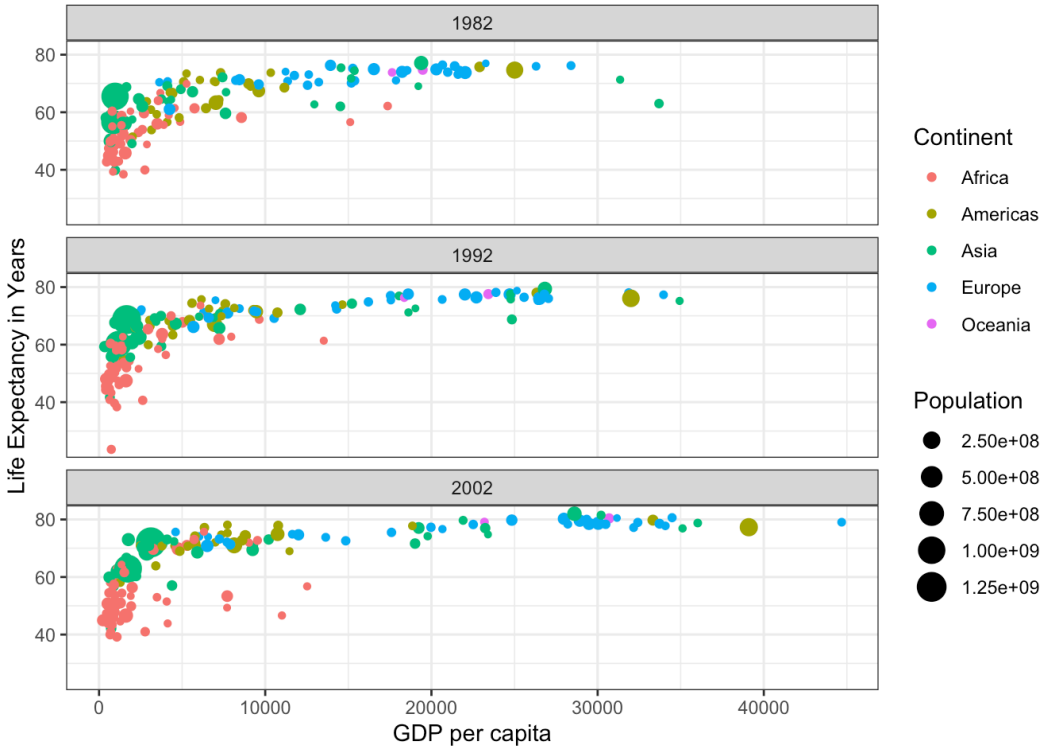
```
library(gapminder)
```

The function filter() will be helpful for completing these question. Though I've referenced it in the class notes a couple of times, we will not fully cover this function until January 23rd. If you want an early start on the homework, see Section 3.2.1 of the book for instructions on how to use filter().

Question 2(a) [10 Points]

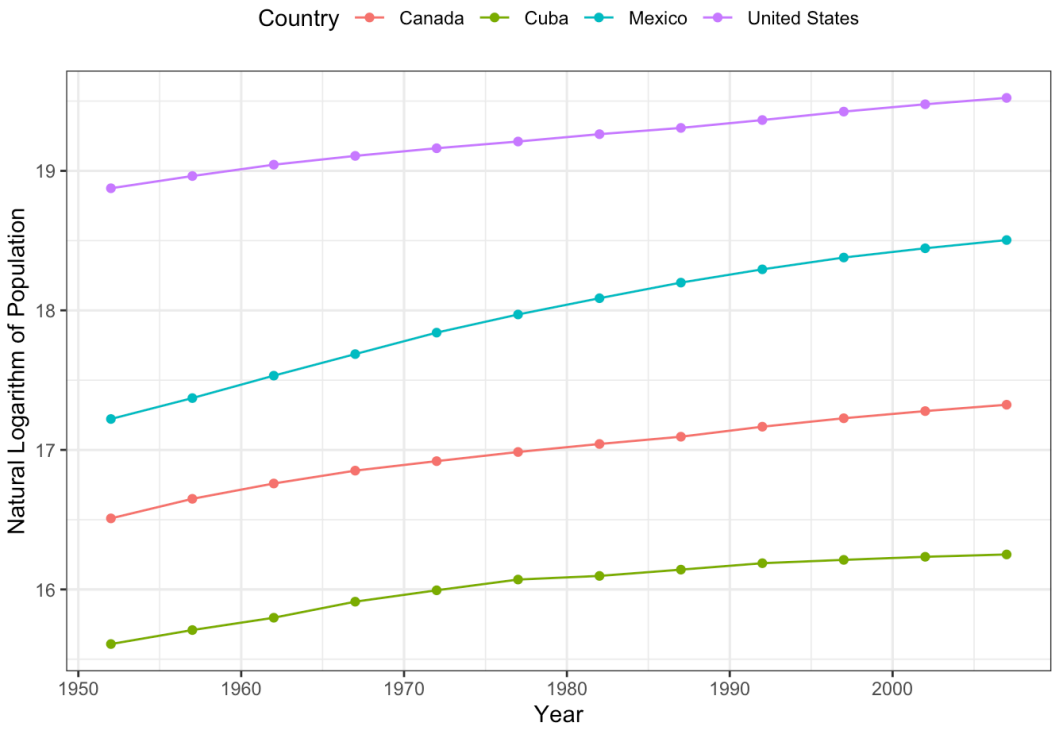
Provide code that recreates the following plot from the gapminder

dataset.



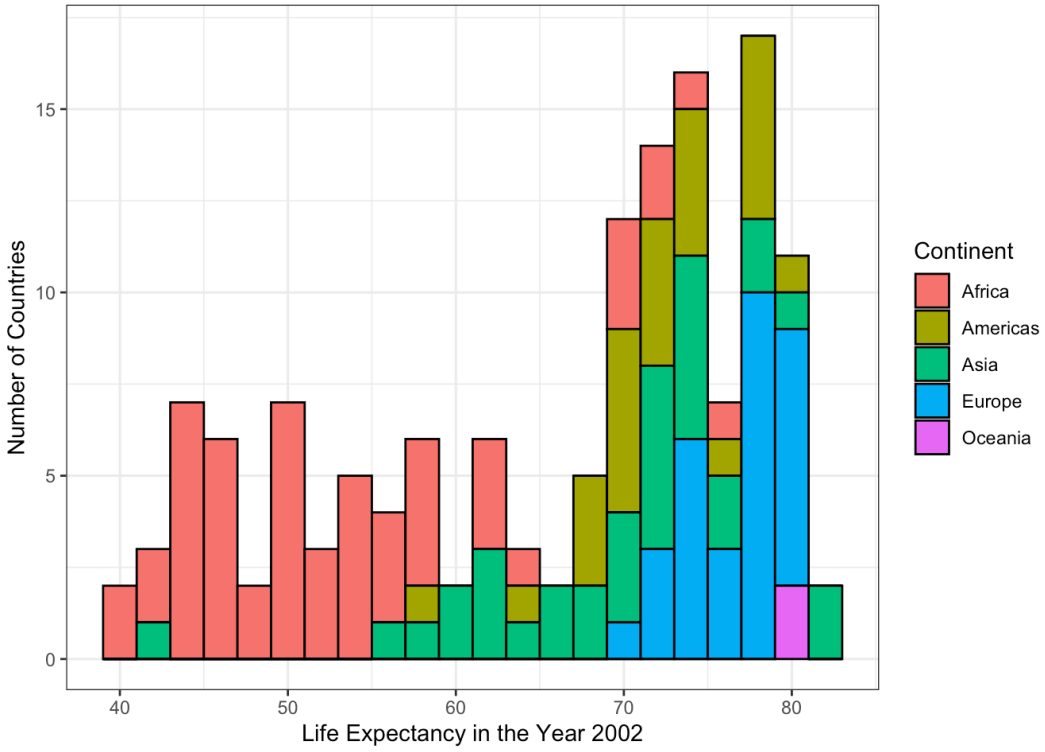
Question 2(b) [10 Points]

Write code to recreate the following plot.



Question 2(c) [10 Points]

Write code to recreate the following plot. Note the bin width is 2.



Question 2(d) [20 Points]

Create three more graphics that visualize the same relationship as in (c).

More specifically, I am looking for three plots that demonstrate the how the life expectancy across countries in 2002 differed by continent. Hint: Consider using different geoms, facets, bar position, colors, etc.

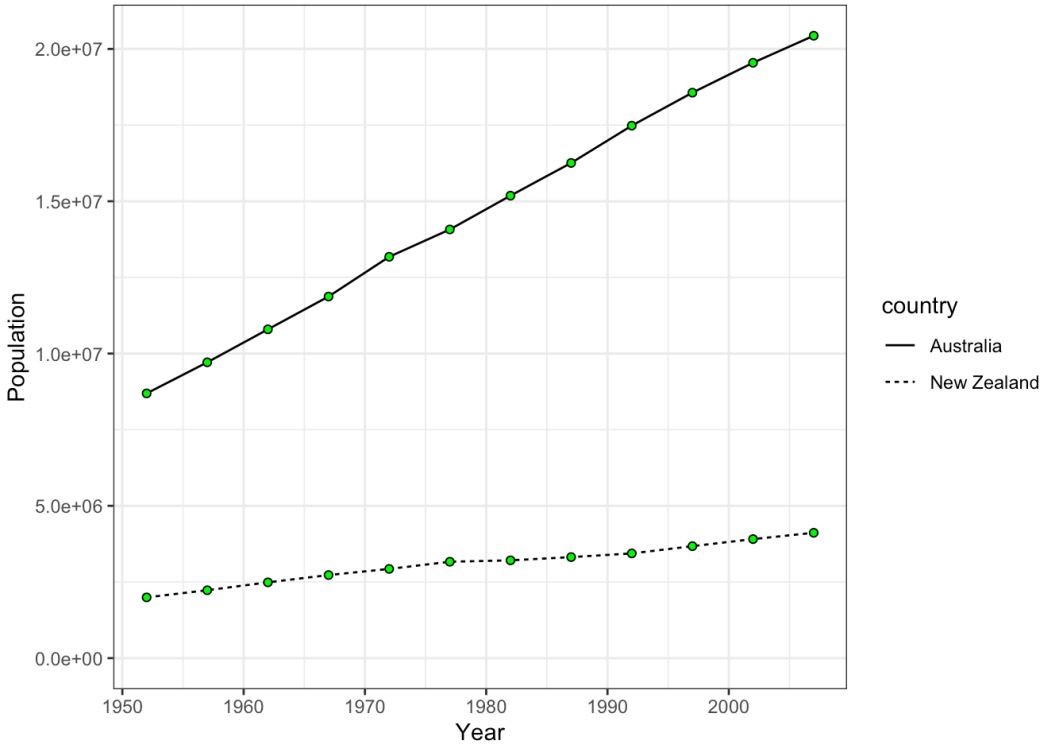
Out of your three, explain which one you feel is the best presentation of the data. Provide a short explanation of your reasoning.

Question 3 [20 Points]

In this question, I will present you with a graphic plot, and code that creates a graphic that doesn't quite accomplish the goal. Your objective is to modify the incorrect code until it creates the desired plot.

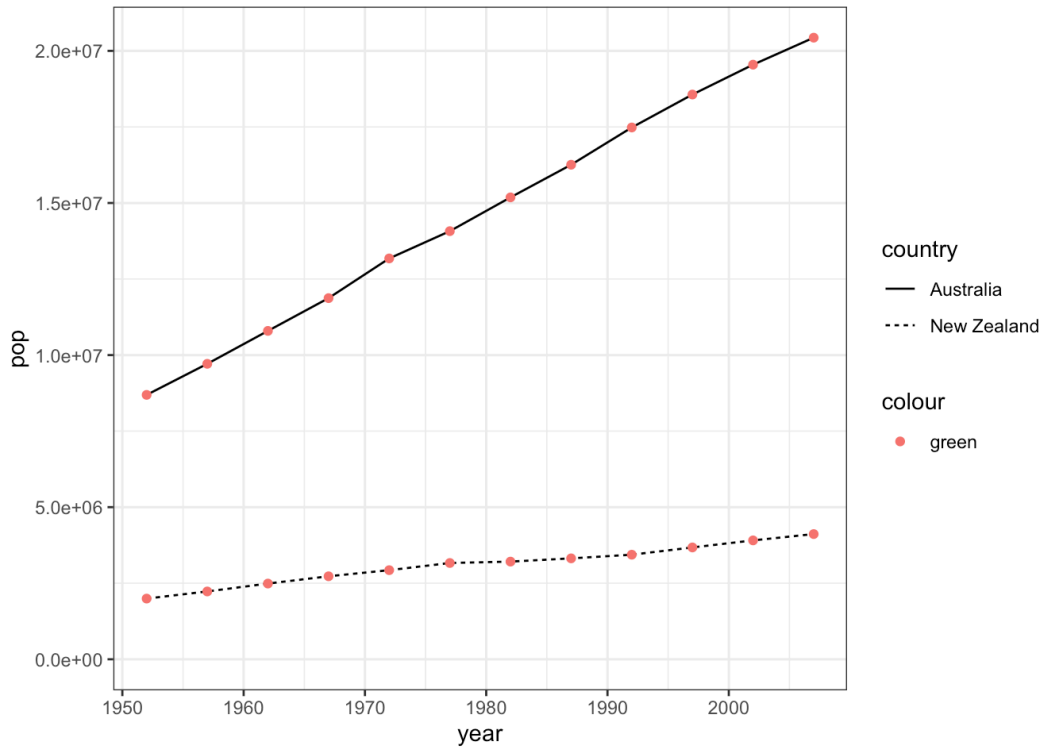
Question 3(a) [5 Points]

Target Plot:



Provided Code and Plot:

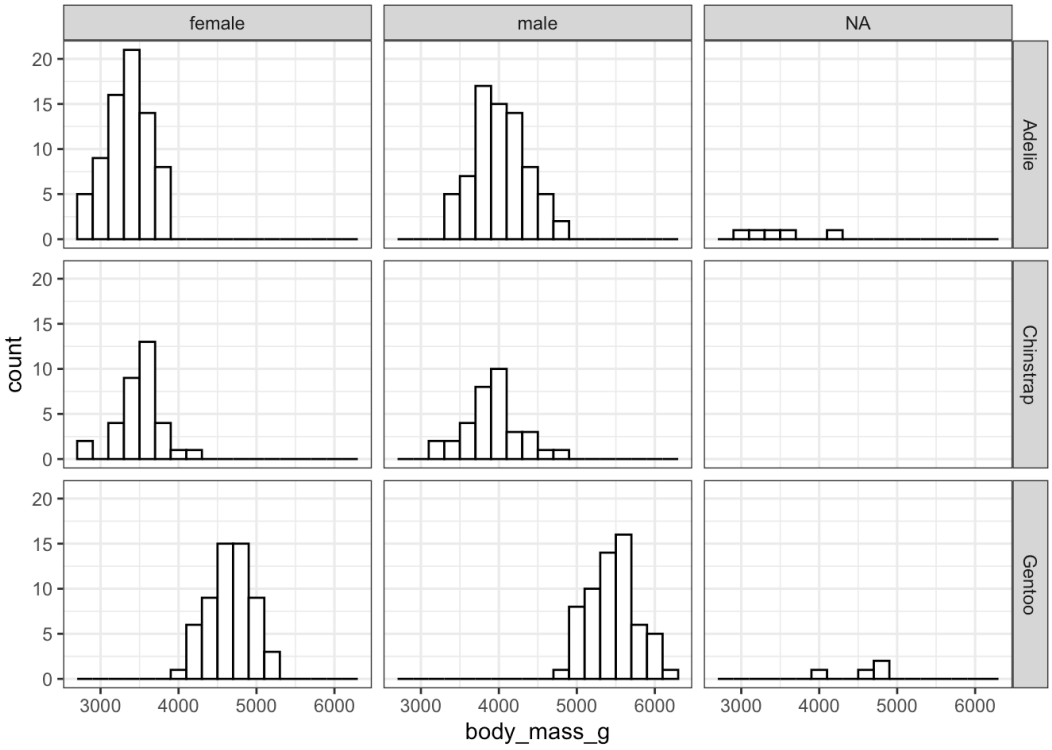
```
gapminder |>
  filter(continent == "Oceania") %>%
  ggplot(aes(x = year, y = pop, linetype = country)) +
  geom_line() +
  geom_point(aes(color = "green")) +
  lims(y = c(0, NA))
```



Question 3(b) [5 Points]

Target Plot:

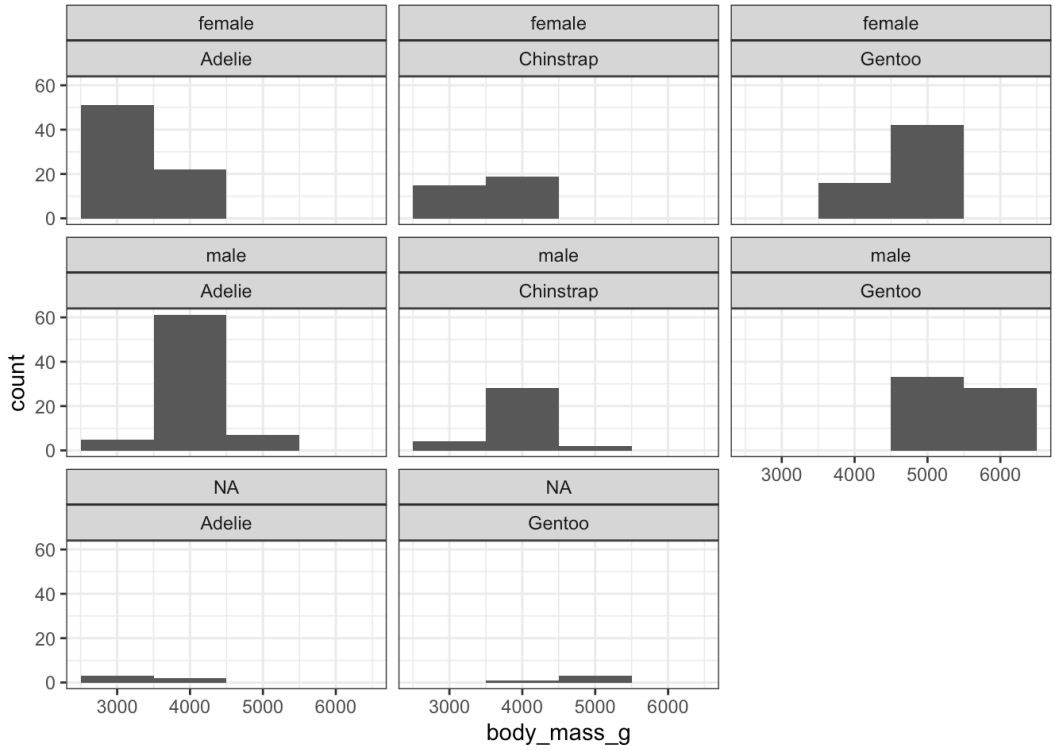
```
Warning: Removed 2 rows containing non-finite values
(`stat_bin()`).
```



Provided Code and Plot:

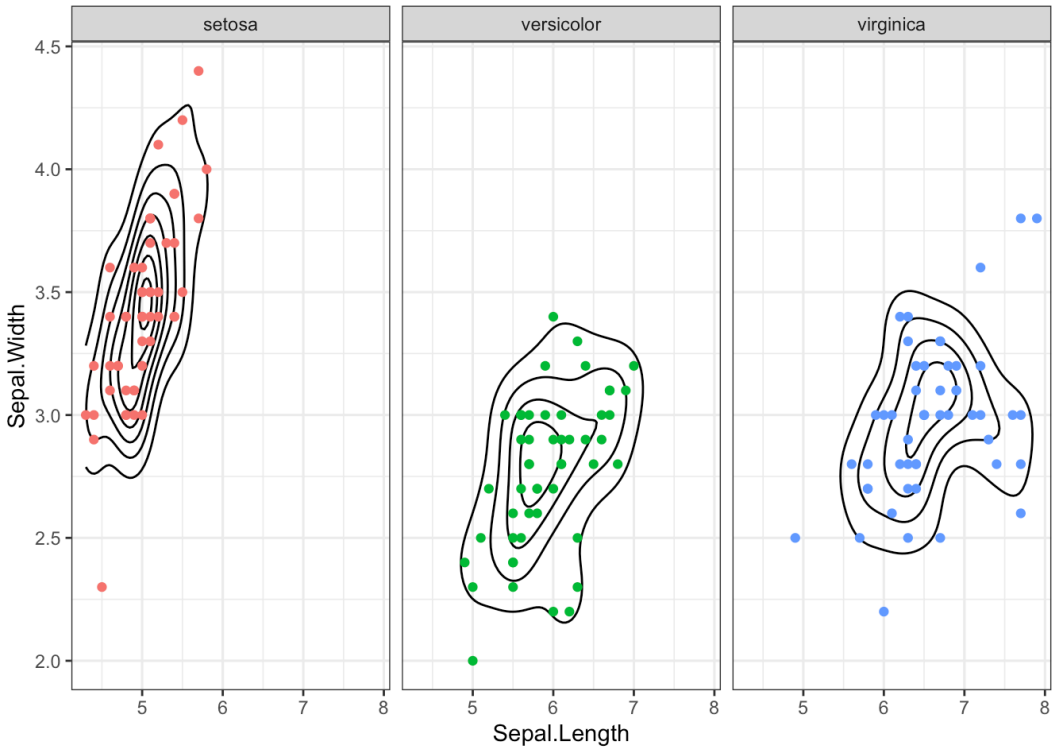
```
library(palmerpenguins)
penguins |>
  ggplot(aes(x = body_mass_g)) +
  geom_histogram(binwidth = 1000) +
  facet_wrap(~sex + species)
```

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



Question 3(c) [5 Points]

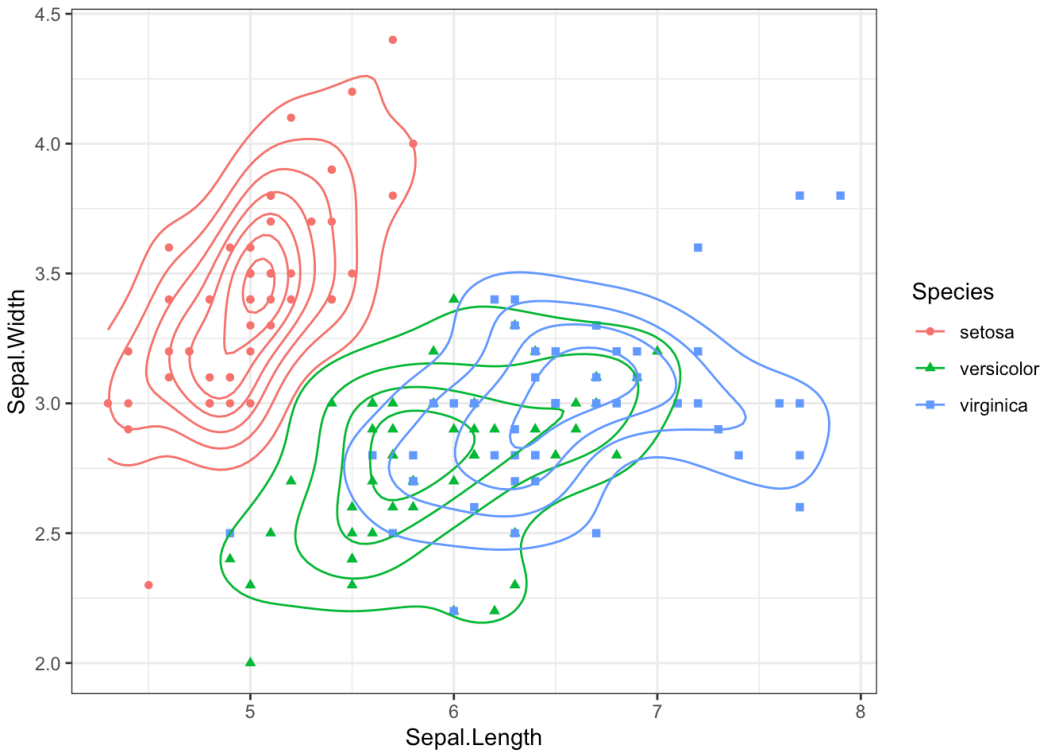
Target plot:



Provided Code and Plot:

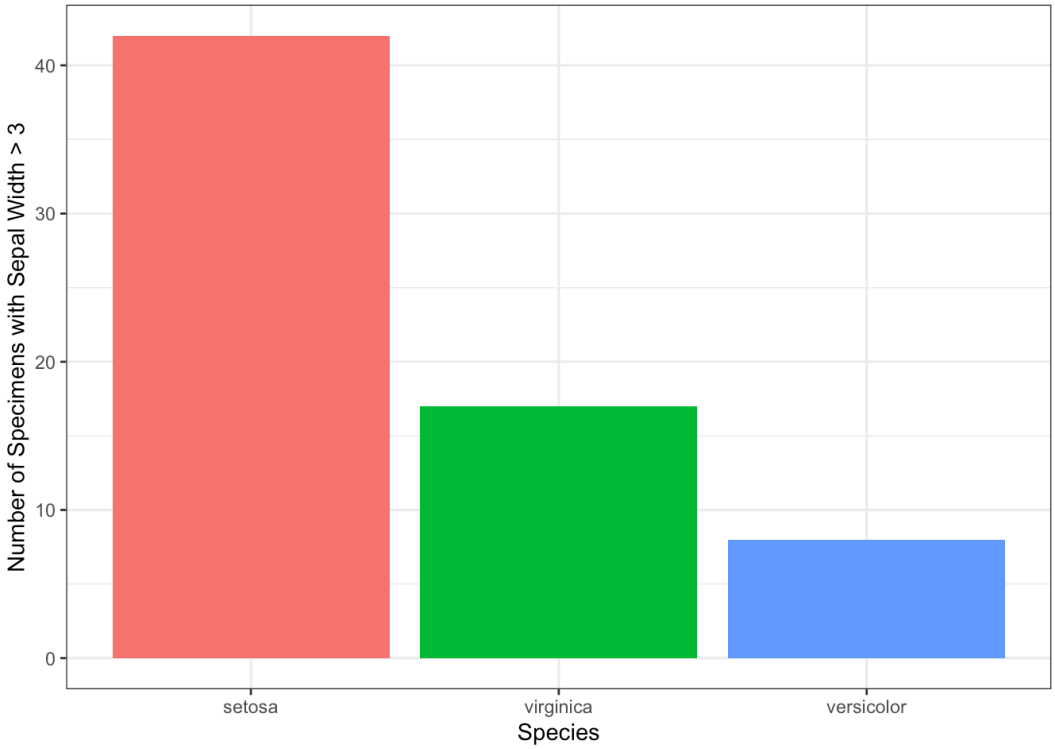
```
set.seed(1)
iris  |>
```

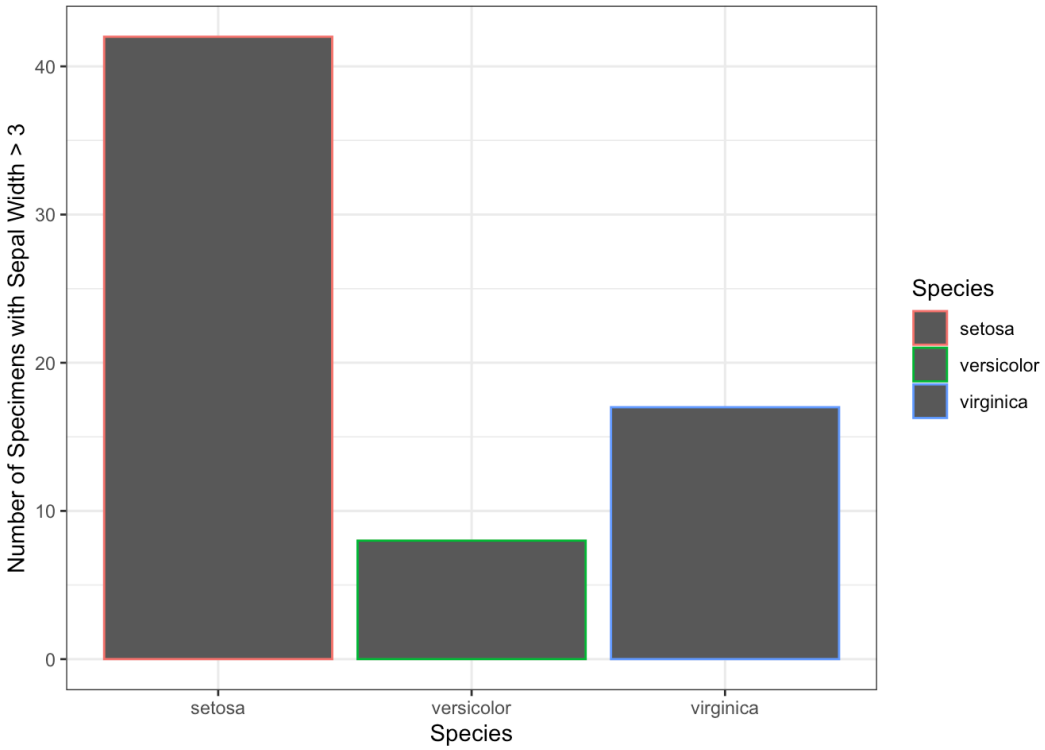
```
ggplot(aes(x = Sepal.Length,
           y = Sepal.Width,
           color = Species,
           shape = Species)) +
geom_point() +
geom_density2d()
```



Question 3(d) [5 Points]

Target plot:



Provided Code and Plot:

```
iris |>
  filter(Sepal.Width > 3) |>
  ggplot(aes(x = Species, color = Species)) +
  geom_bar() +
  labs(y = "Number of Specimens with Sepal Width > 3")
```