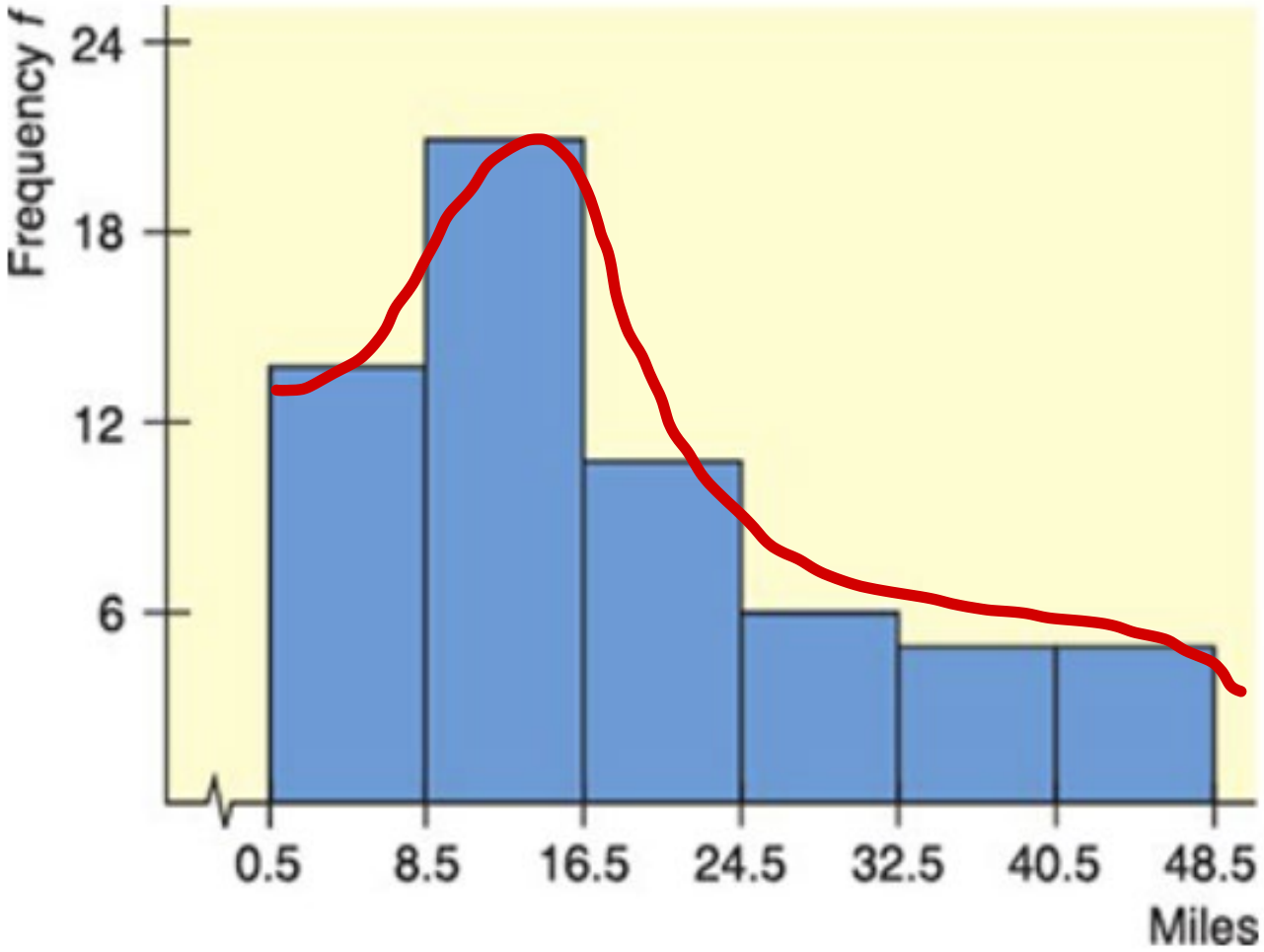
Data Visualization
Part 1

One variable plots

Histograms & Density

numeric variable

Commuting_Dist
1
5
5
6
7
4
8
7
6
5
5
6
7
…



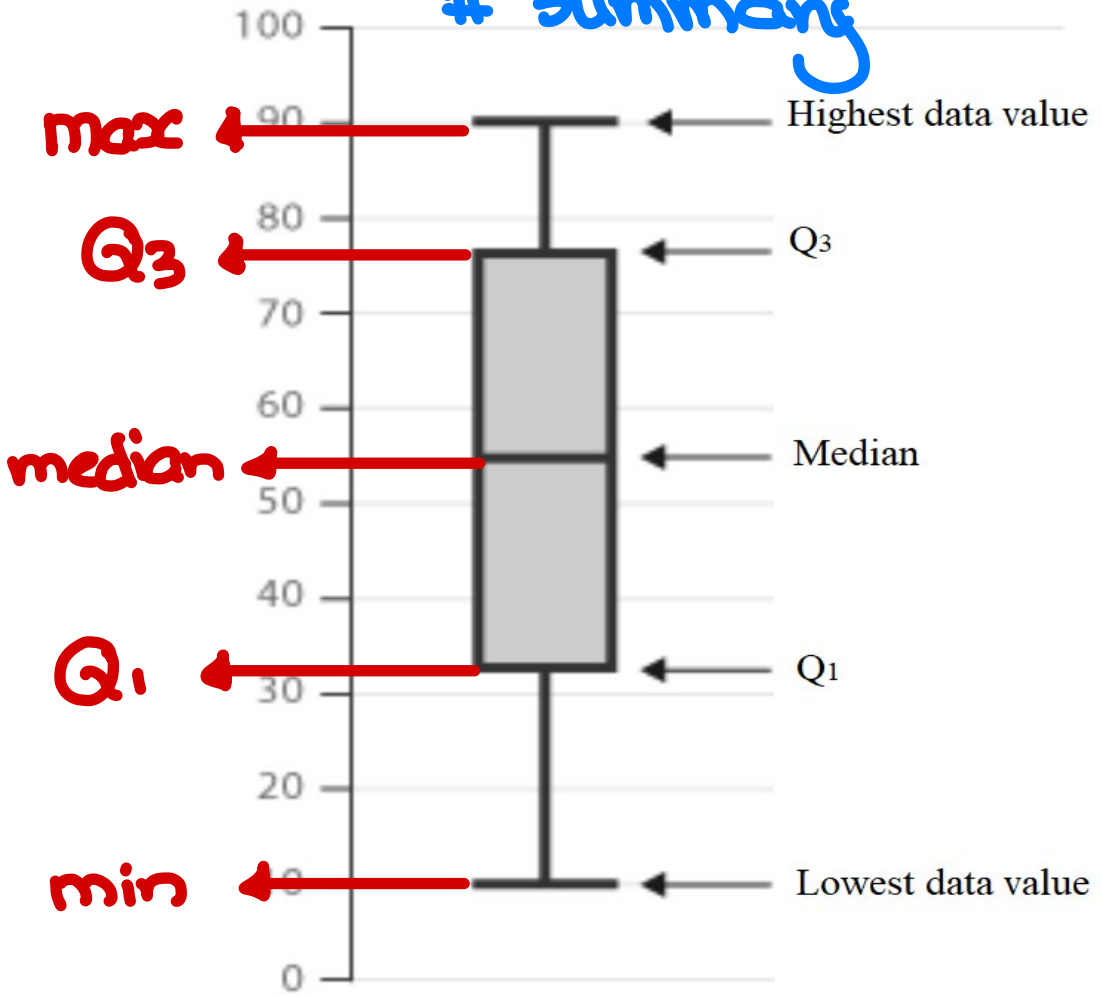
Boxplots

5 # Summary
min, Q1, median, Q3, max

numeric variable

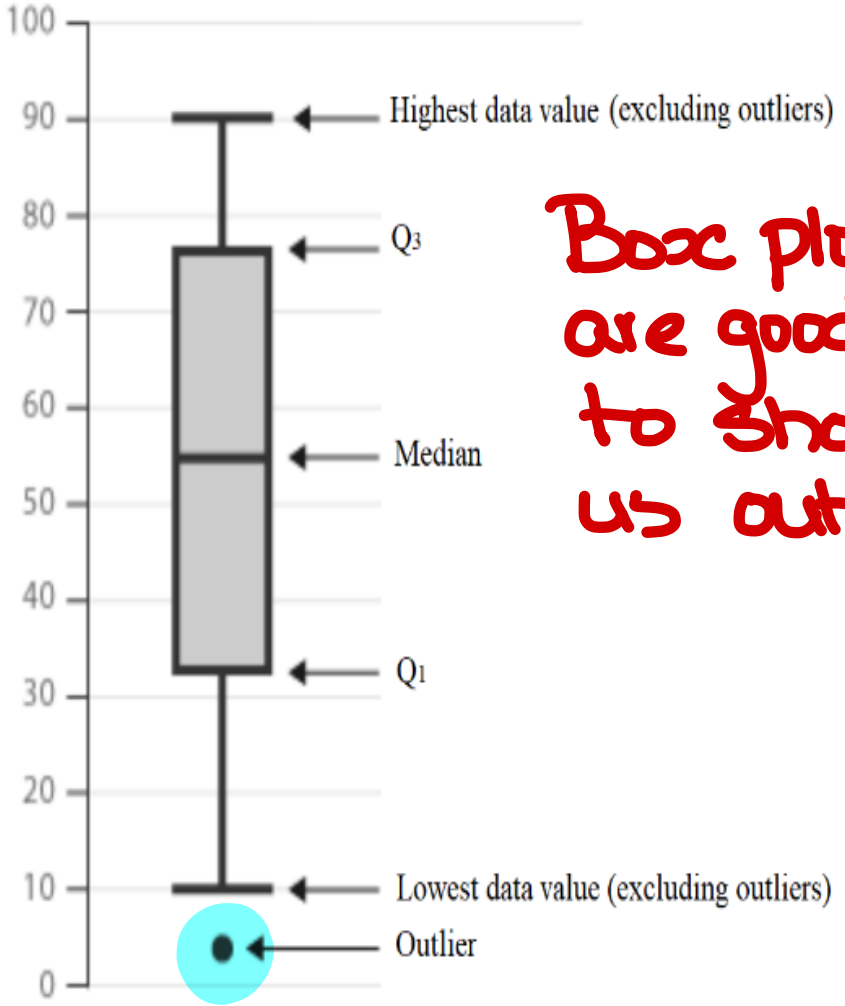
Time_Exam
90
80
75
80
50
55
45
40
10
20
25
30
35
40
…

Boxplot is a plot of the 5 # summary



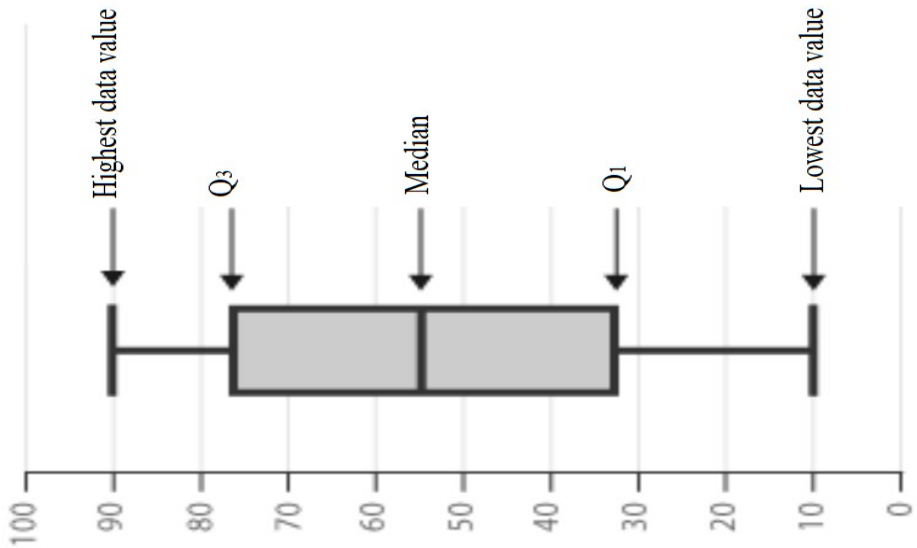
Other ways of representing boxplots:

Vertical Boxplot



Box plots are good to show us outliers

Horizontal Boxplot

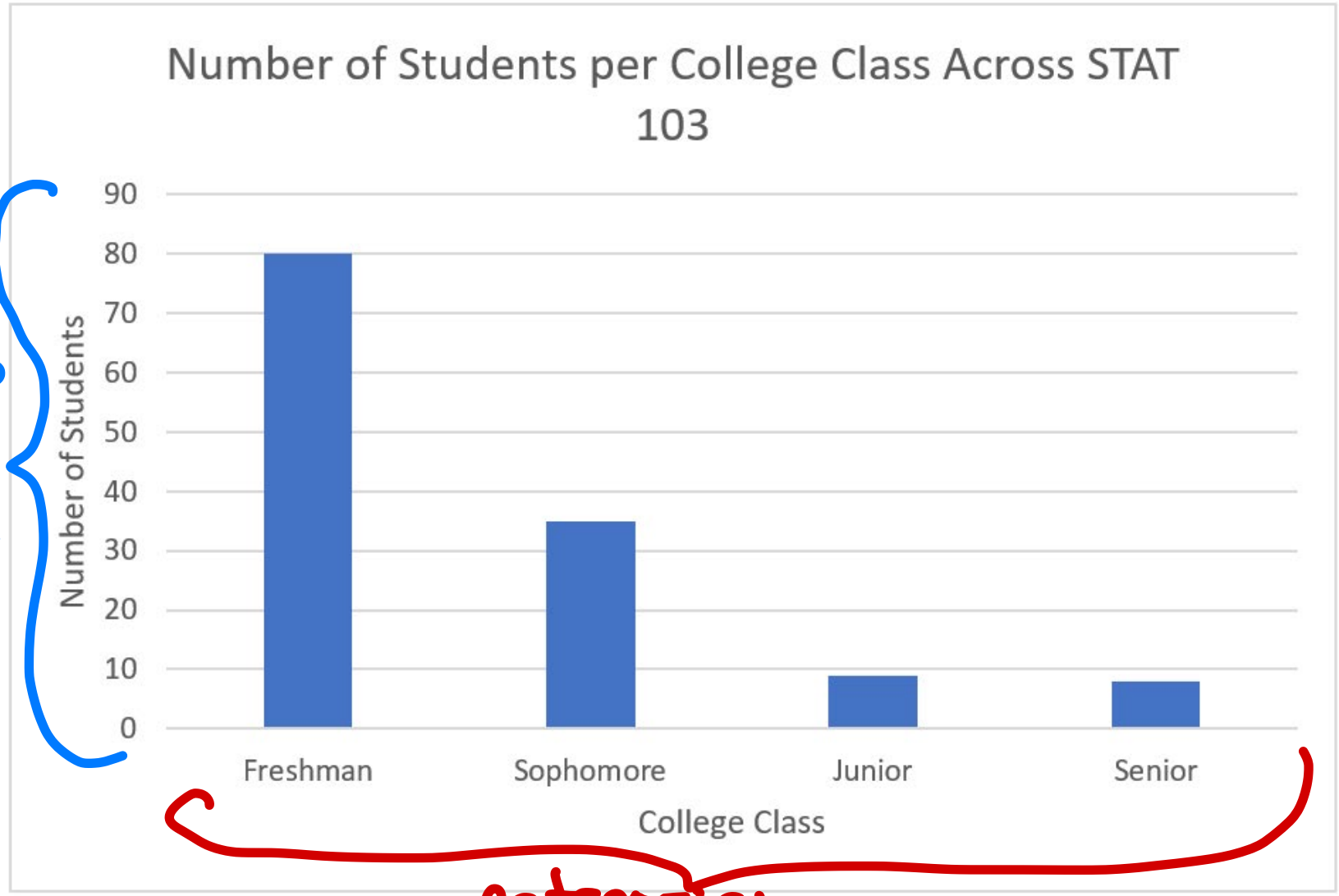


Categorical Variable

Bar Graphs

Grad_Class
Freshman
Freshman
Freshman
Junior
Senior
Sophomore
Freshman
Sophomore
Freshman
Freshman
Junior
Sophomore
Freshman
Freshman
Sophomore
…

Counts by default



Categories

`ggplot()` Function

- **Data:** The dataset you're working with.
- **Aesthetics Mapping (aes):** How data variables map to plot aesthetics like position, color, shape, etc.

```
x = variable1
y = variable2
color = variable3
fill = variable4
shape = variable5
```

- **Geometric Objects (geom):** The visual elements to represent the data (points, lines, bars, etc.).

```
geom_histogram()
geom_density()
geom_boxplot()
geom_line()
geom_bar()
geom_point()
```

- **Facets (facet_wrap or facet_grid):** Splitting data into subplots based on a variable.

```
facet_wrap(~ Variable1)
facet_grid(Variable1 ~ Variable2)
```

- **Theme:** Controlling the overall appearance of the plot.

```
theme_gray()
theme_bw()
theme_minimal()
theme_void()
```

- **Add ons**

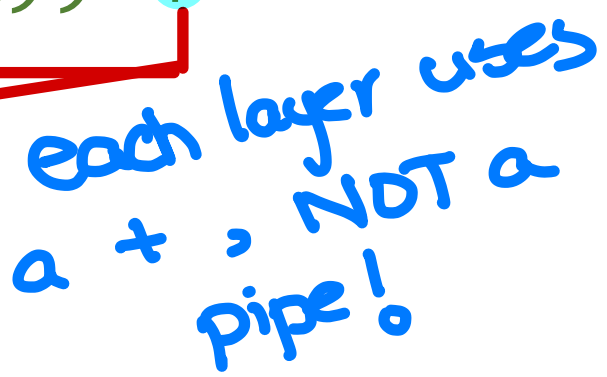
```
coord_flip()
geom_smooth(method = "lm", se = FALSE)
geom_smooth(method = "loess", se = TRUE)
scale_fill_manual(values = c("red", "green", "blue"))
scale_color_manual(values = c("red", "green", "blue"))
labs()
```

flips your coordinates

ggplot() Function

Command Illustration

```
Plot_Name <- name_daframe %>%
     ggplot(aes(x = column_name)) +
     geom_OBJECT()+
     facet()+
     theme()
```



Type : ?SAT_2010
in your console
to learn about the dataset

Data

	state	expenditure	pupil_teacher_ratio	salary	read	math	write	total	sat_pct	salary_level
1	Alabama	10	15.3	49948	556	550	544	1650	8	Low
2	Alaska	17	16.2	62654	518	515	491	1524	52	High
3	Arizona	9	21.4	49298	519	525	500	1544	28	Low
4	Arkansas	10	14.1	49033	566	566	552	1684	5	Low
5	California	10	24.1	71611	501	516	500	1517	53	High
6	Colorado	10	17.4	51660	568	572	555	1695	19	Low
7	Connecticut	16	13.1	67565	509	514	513	1536	87	High
8	Delaware	13	14.5	59932	493	495	481	1469	74	Medium
9	Florida	9	15.1	49042	496	498	479	1473	64	Low
10	Georgia	10	14.9	55766	488	490	475	1453	80	Medium
11	Hawaii	13	15.8	57814	483	505	470	1458	64	Medium
12	Idaho	7	17.6	48596	543	541	517	1601	20	Low

Histograms

Figure 1: Create a Histogram on the Average Math SAT Score.

```
fig_1 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_histogram()
fig_1
```

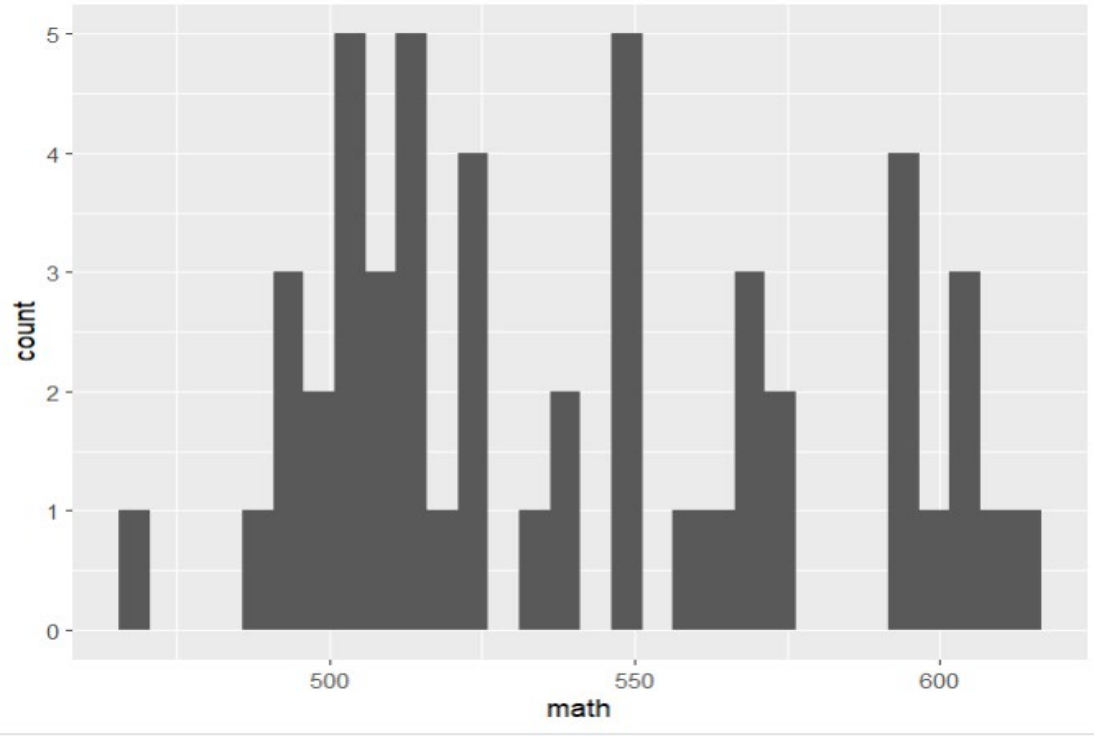


Figure 2: Create a Histogram on the Average Math SAT Score, but change the break points to start from 450, end in 625, and increase by 25

```
fig_2 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_histogram(breaks =
c(450,475,500,525,550,575,600,625))

fig_2
```

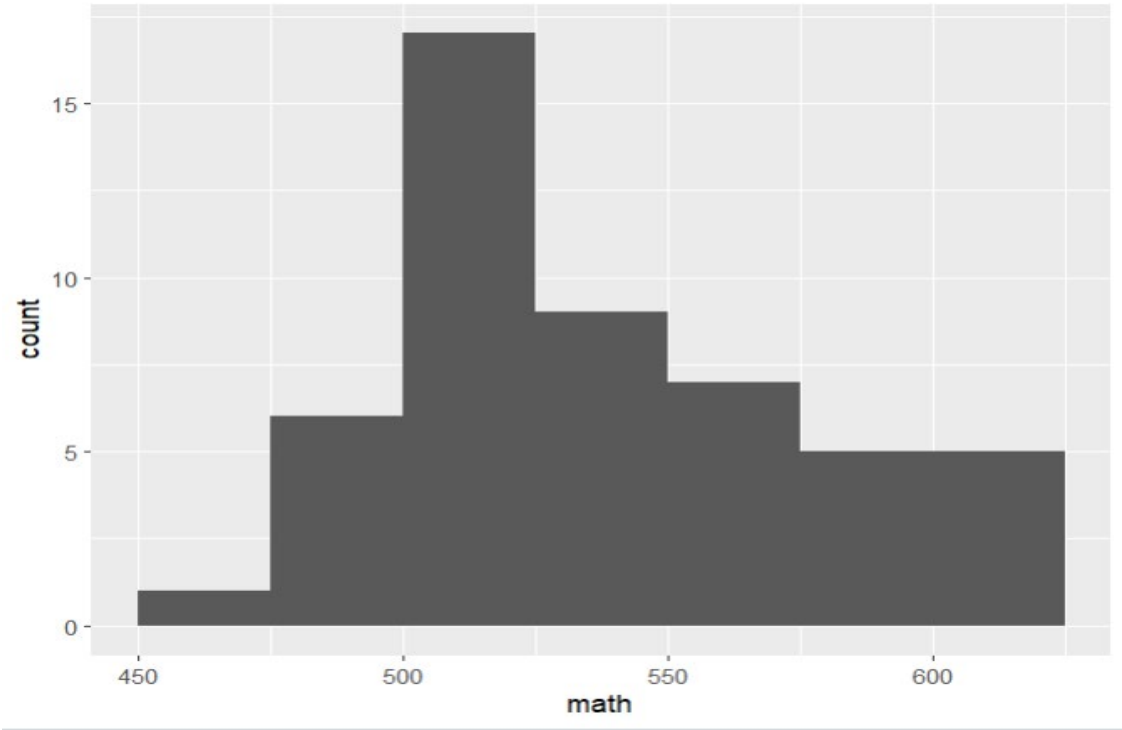
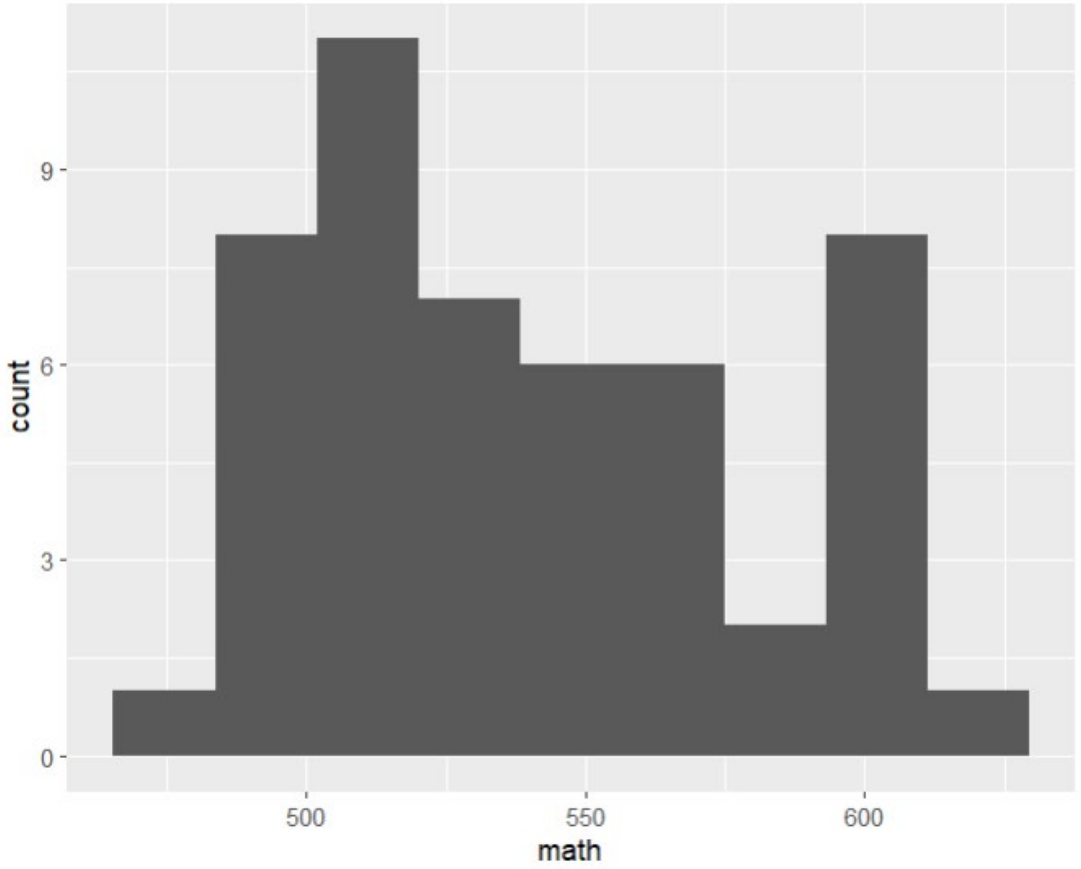


Figure 3: Create a Histogram on the Average Math SAT Score, but choose the bins to be 9.

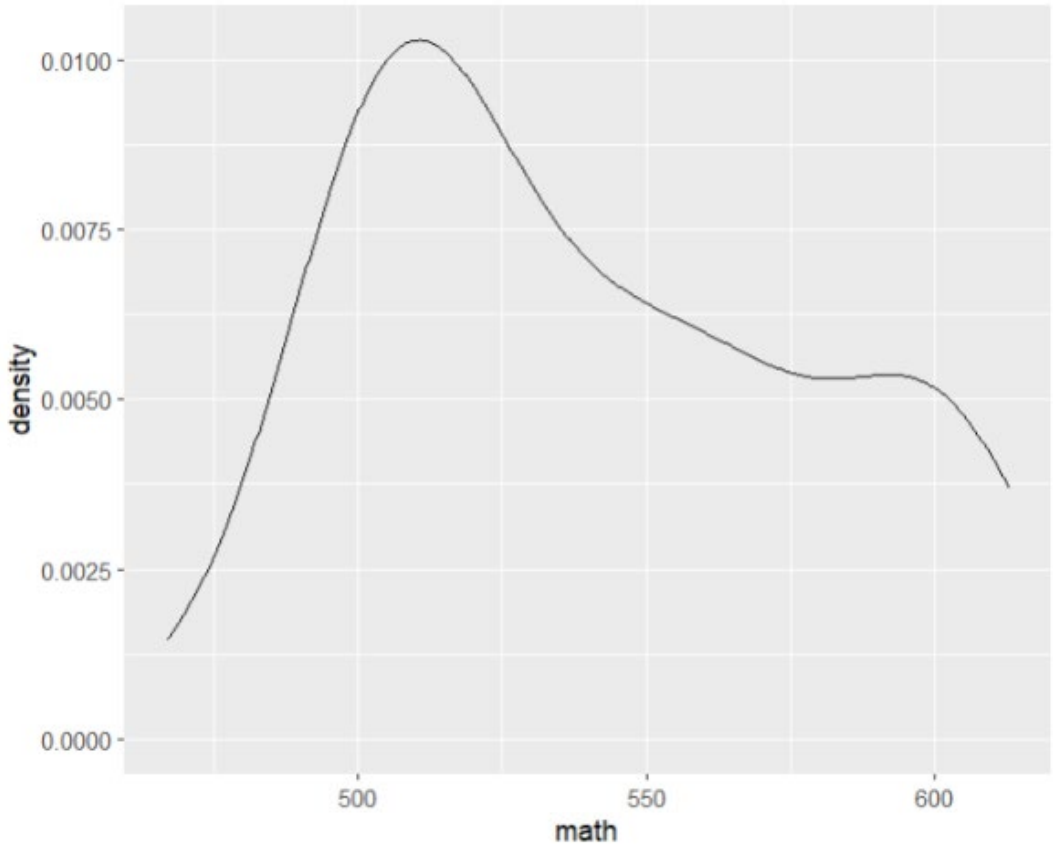
```
fig_3 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_histogram(bins = 9)
fig_3
```



Density Plots

Figure 4: Create a density plot on the Average Math SAT Score

```
fig_4 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_density()
fig_4
```



bandwidth → bw

Figure 5: Create a density plot on the Average Math SAT Score. Change the bandwith to 10.

Note: Changing the bandwidth to a smaller number makes it more jagged.

```
fig_5 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_density(bw = 10)
fig_5
```

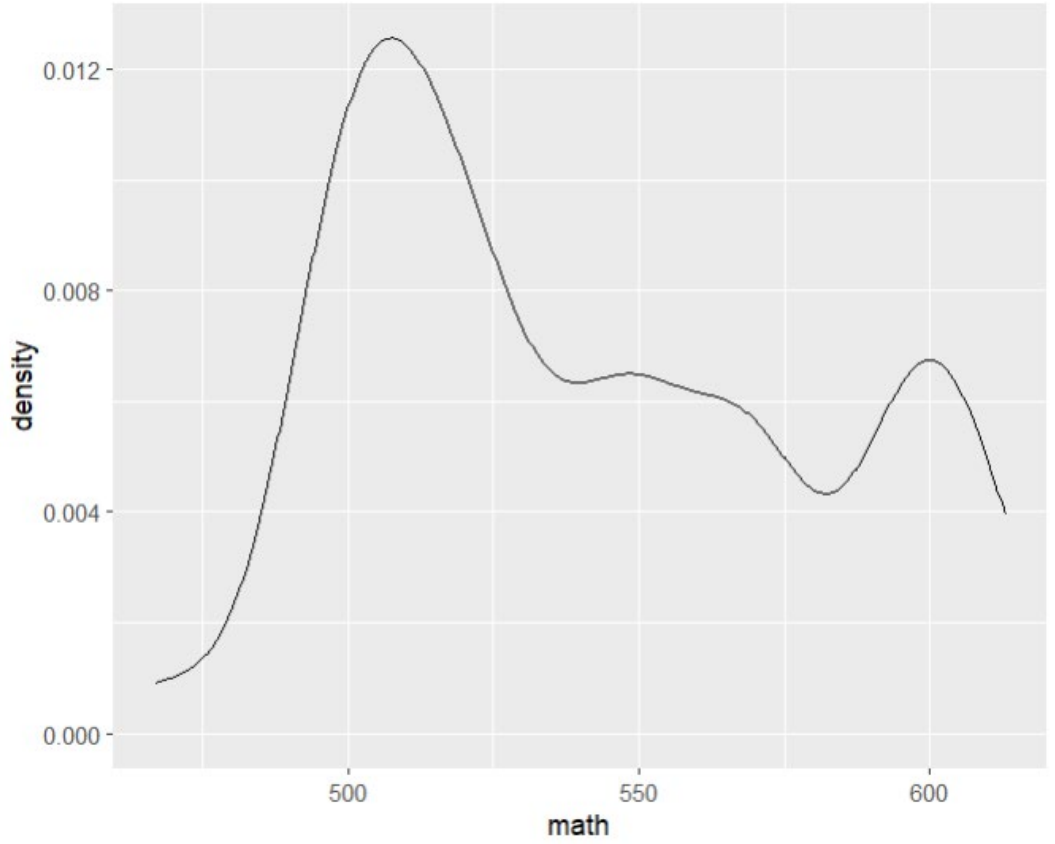
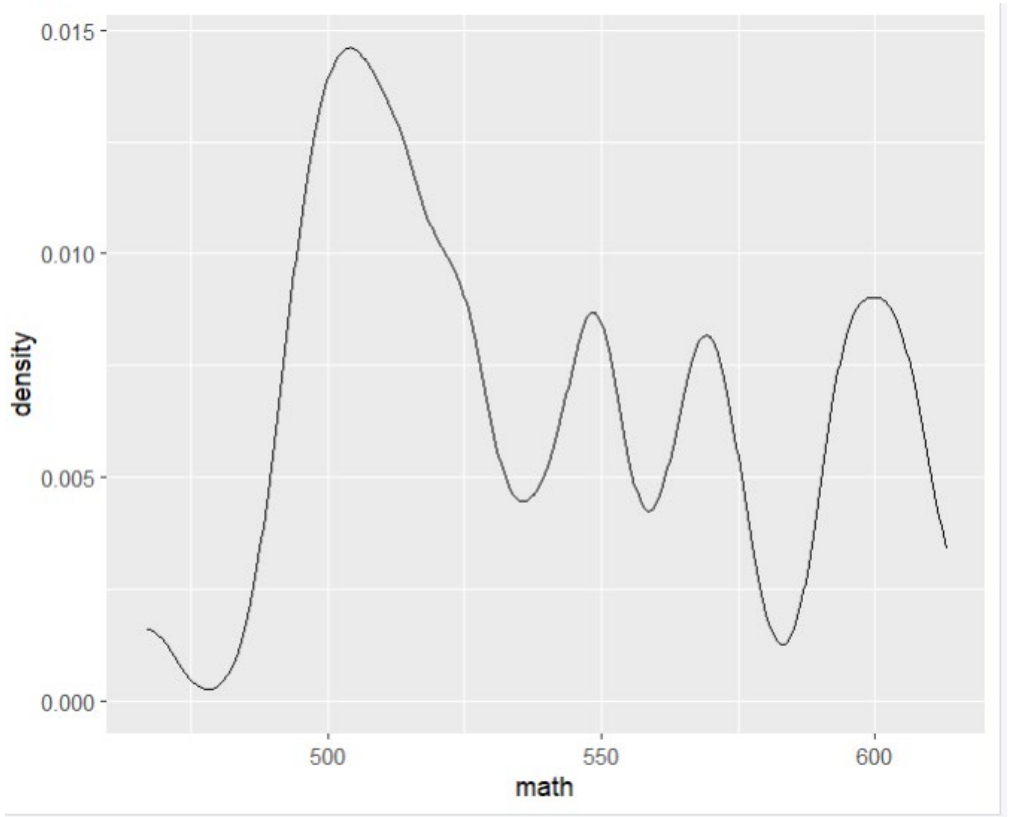


Figure 6: Create a density plot on the Average Math SAT Score. Change the bandwith to 5.

```
fig_6 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_density(bw = 5)
fig_6
```



Boxplots

Figure 7: Create a box plot on the Average Math SAT Score.

```
fig_7 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_boxplot()
fig_7
```

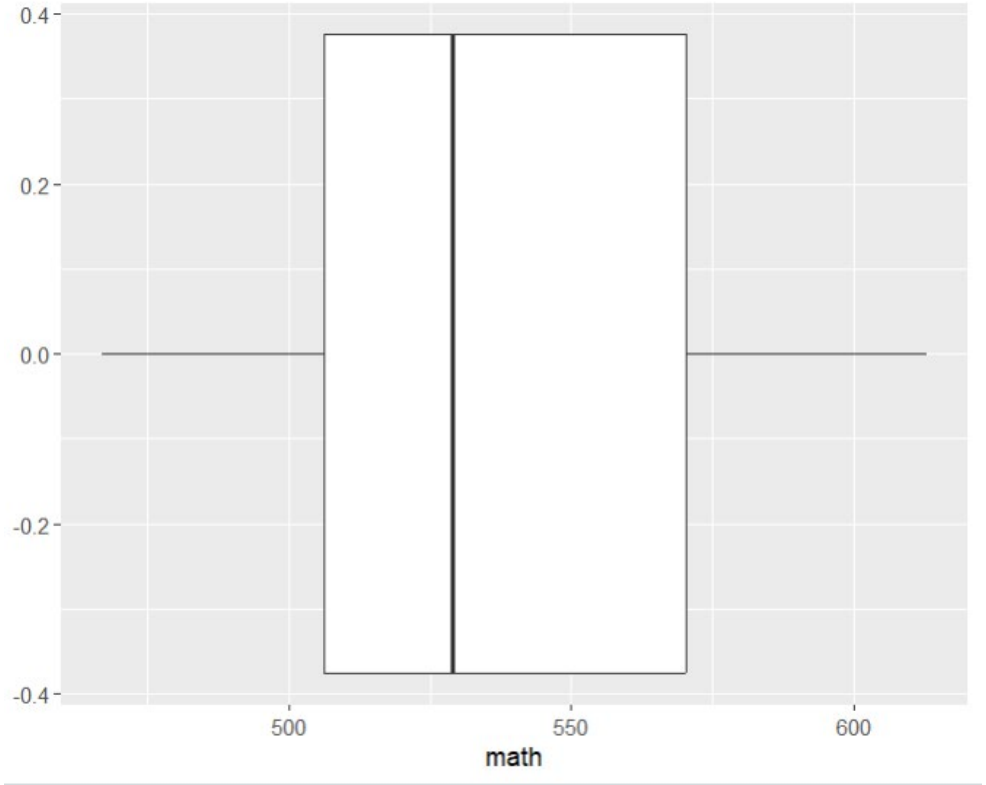


Figure 8: Create a box plot on the Average Math SAT Score. Make the boxplot vertical.

```
fig_8 <- SAT_2010 %>%
  ggplot(aes(x = math)) +
  geom_boxplot() +
  coord_flip()
fig_8
```

flips the coordinates

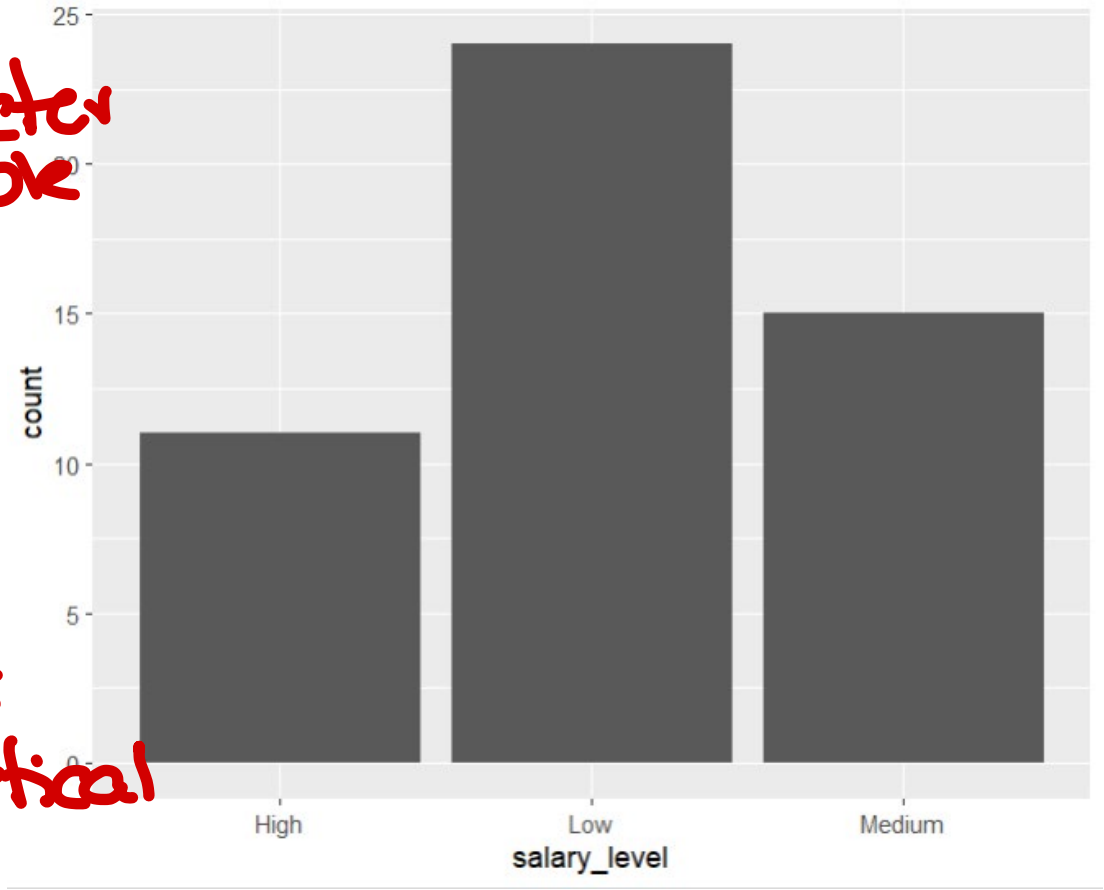
class(SAT_2010$salary_level) = character.

Bar Plots

Figure 9: Create a bar graph on the Salary Level (the new variable you created).

```
fig_9 <- SAT_2010 %>%
  ggplot(aes(x = salary_level)) +
  geom_bar()
fig_9
```

Character variable



If variable is character, categories in your plot will appear in alphabetical order

Figure 10: Create a bar graph on the Salary Level (the new variable you created). Have the category "Low" appear first, then "Medium", then "High".

```
fig_10 <- SAT_2010 %>%
    mutate(salary_level = factor(salary_level,
    levels = c("Low",  "Medium", "High"))) %>%
        ggplot(aes(x = salary_level)) +
        geom_bar()
fig_10
```

turn it into a factor variable

order you want your bars to show