Lab_05

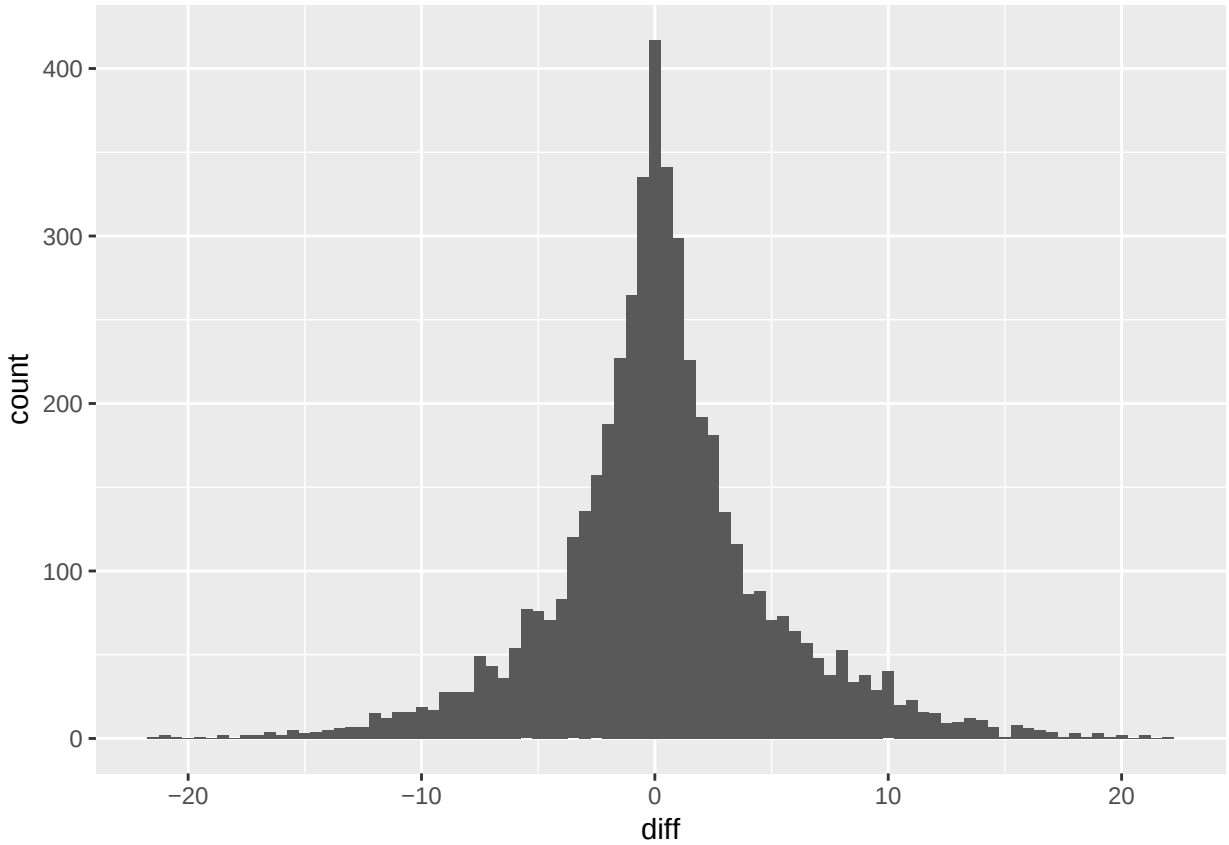
Fantastic Four

Guiding Question

What is the relationship between the resulting spectra after the ionization of partices? What variation occurs for each spectra, spec5 and spec10? What covaritation occurs?

Overall Findings

```
#histogram of difference between spec10 and spec5(for team)
ms%>%
  filter(spec5>0 & spec5<25, spec10>0 & spec10<25)%>%
  #summarize(spec5AVG=mean(spec5), spec10AVG=mean(spec10))%>%
  mutate(diff=spec10-spec5)%>%
  filter(mass<200)%>%
  ggplot(mapping=aes(x=diff))+ geom_histogram(binwidth = 0.5)
```



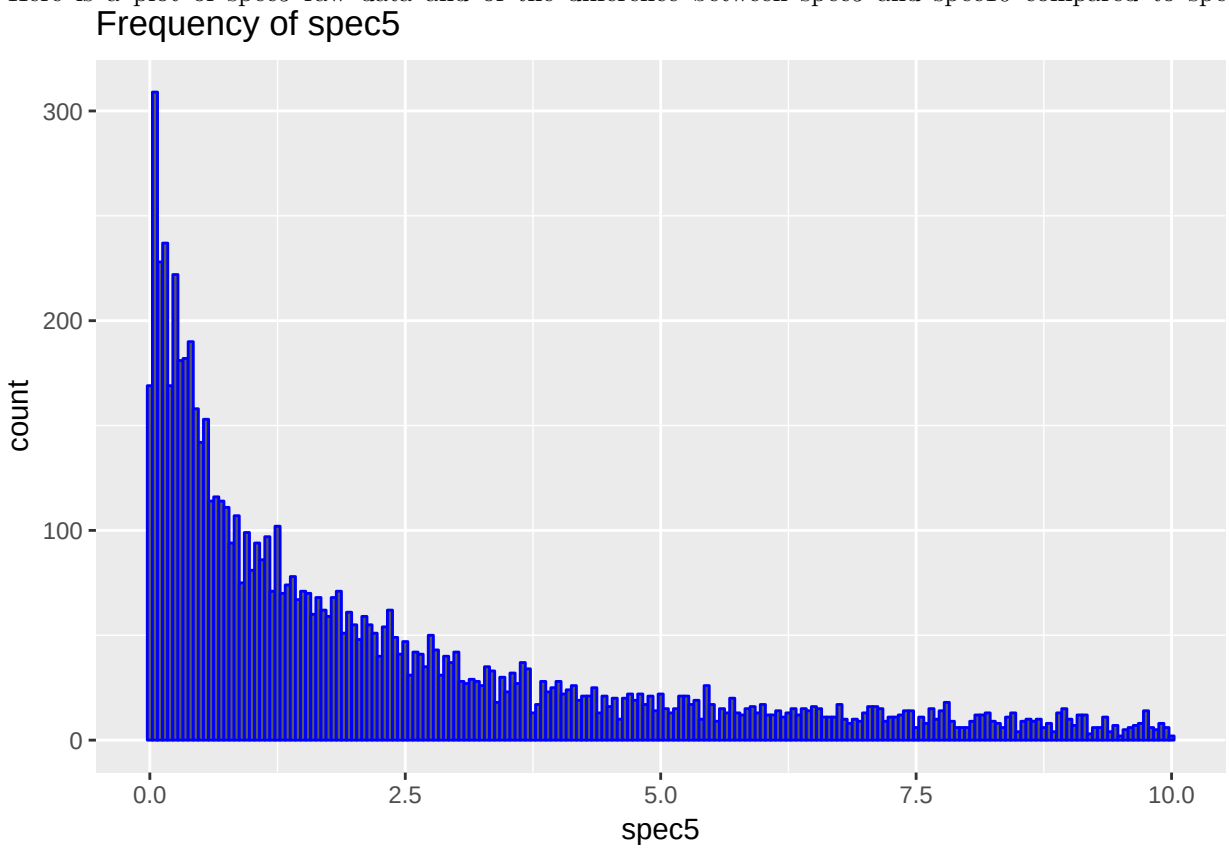
This graph shows the difference between the spec5 and spec10 variables. As seen, the average difference is about zero. This means that spec10 and spec5 are equally produced when ionized. Additionall through individual findings:

- Most frequent occurrences for low mass/charge ratios
- spec5 is slightly more common
- Wide range of resulting spectra values (0,400) and negative values (broad span of data)
- Significant readings for mass/charge around 30 and 50
- Lots of noise (possible detection of other ions)

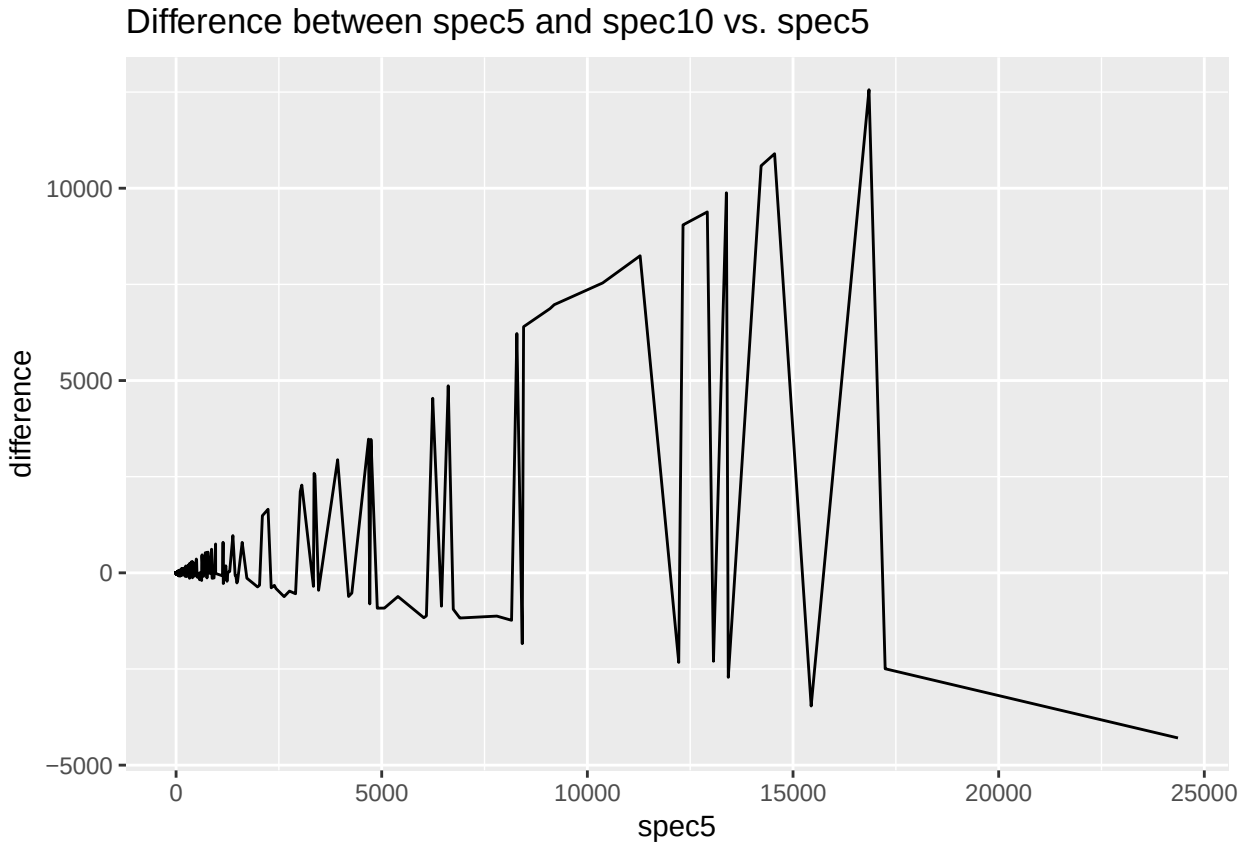
Individual Findings

Lexie Marinelli

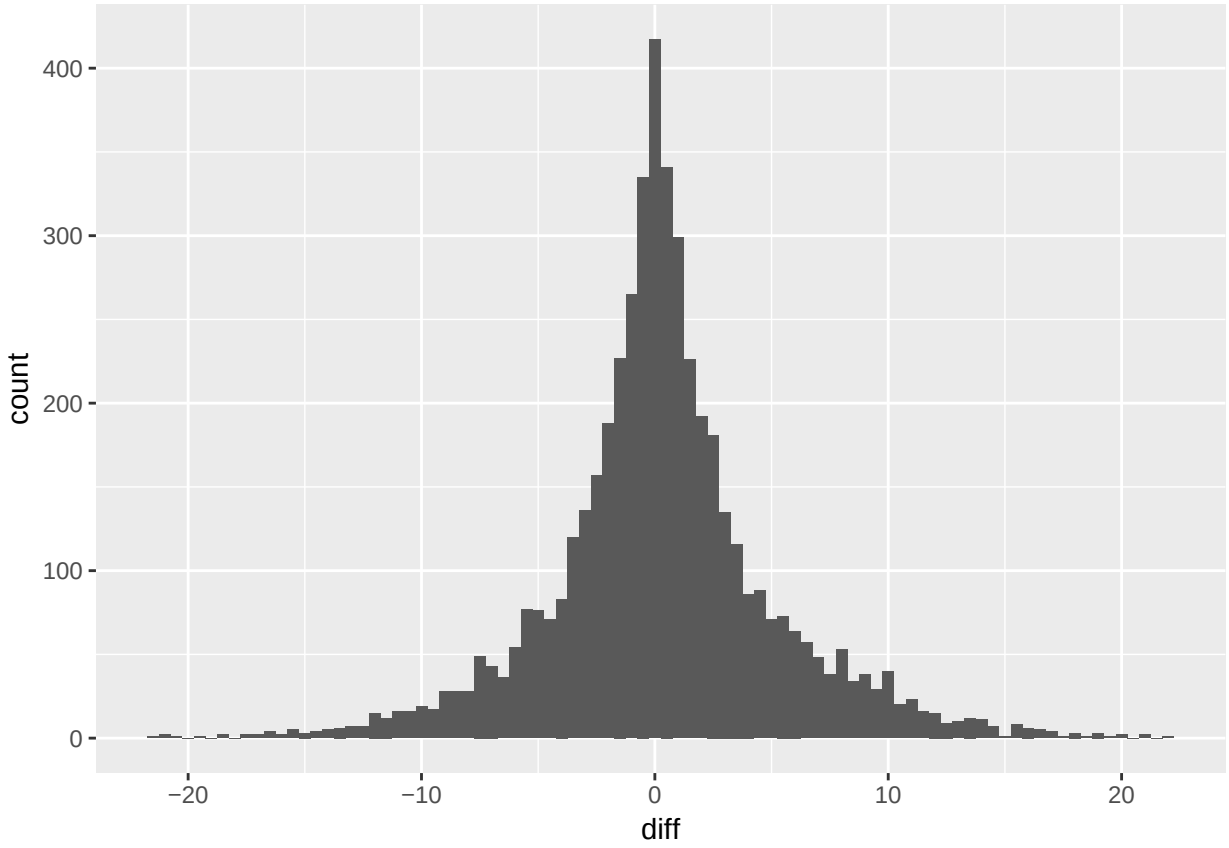
Here is a plot of spec5 raw data and of the difference between spec5 and spec10 compared to spec5.



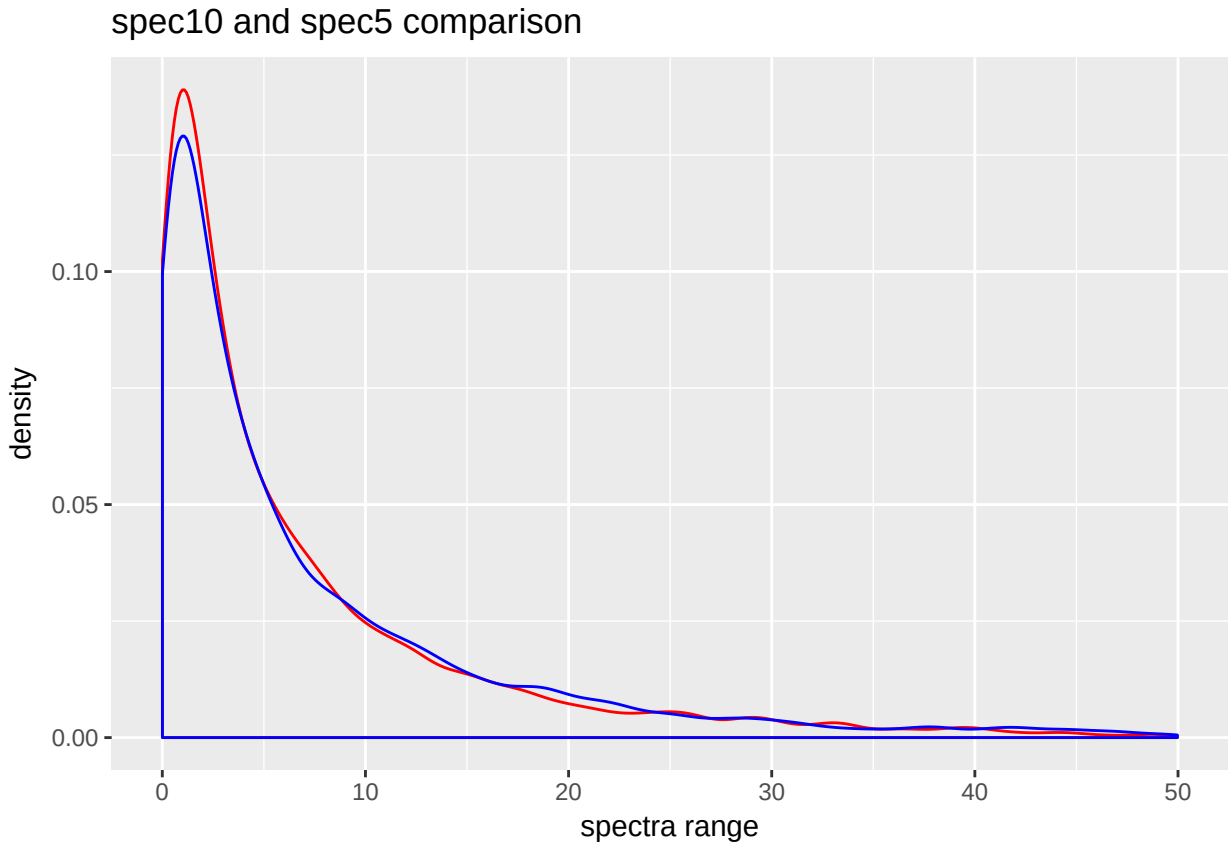
The first graph shows the relationship for spec5 as it gets larger, the occurrence of those values faces an exponential decrease.

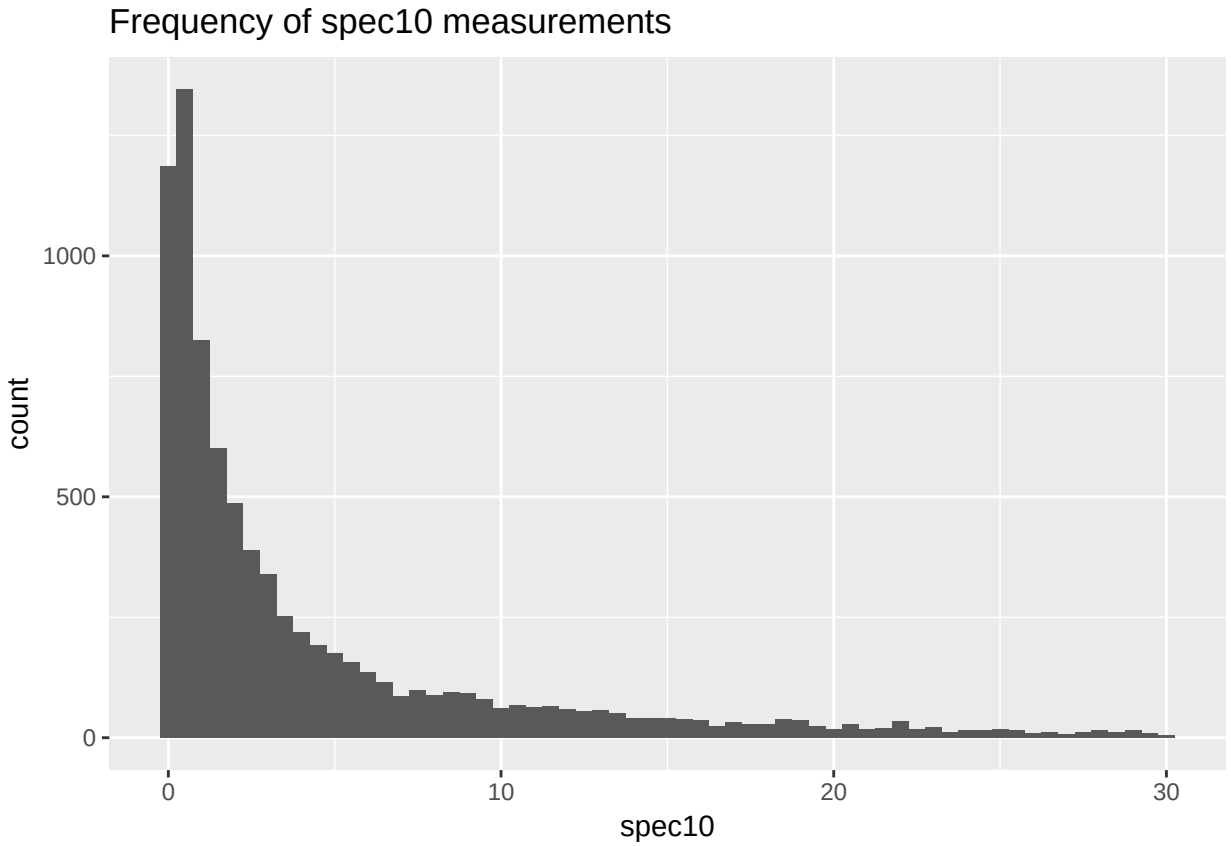


The second graph tells us that as the value for spec5 increases, the difference between the values of spec5 and spec10 increase and oscillates to form a v shape.



The first graph models the differences between the spec5 and spec10. This shows that between the two spectra outputs, are most frequent for smaller masses. At smaller masses there is also a greater difference in the values of spec5 and spec10.





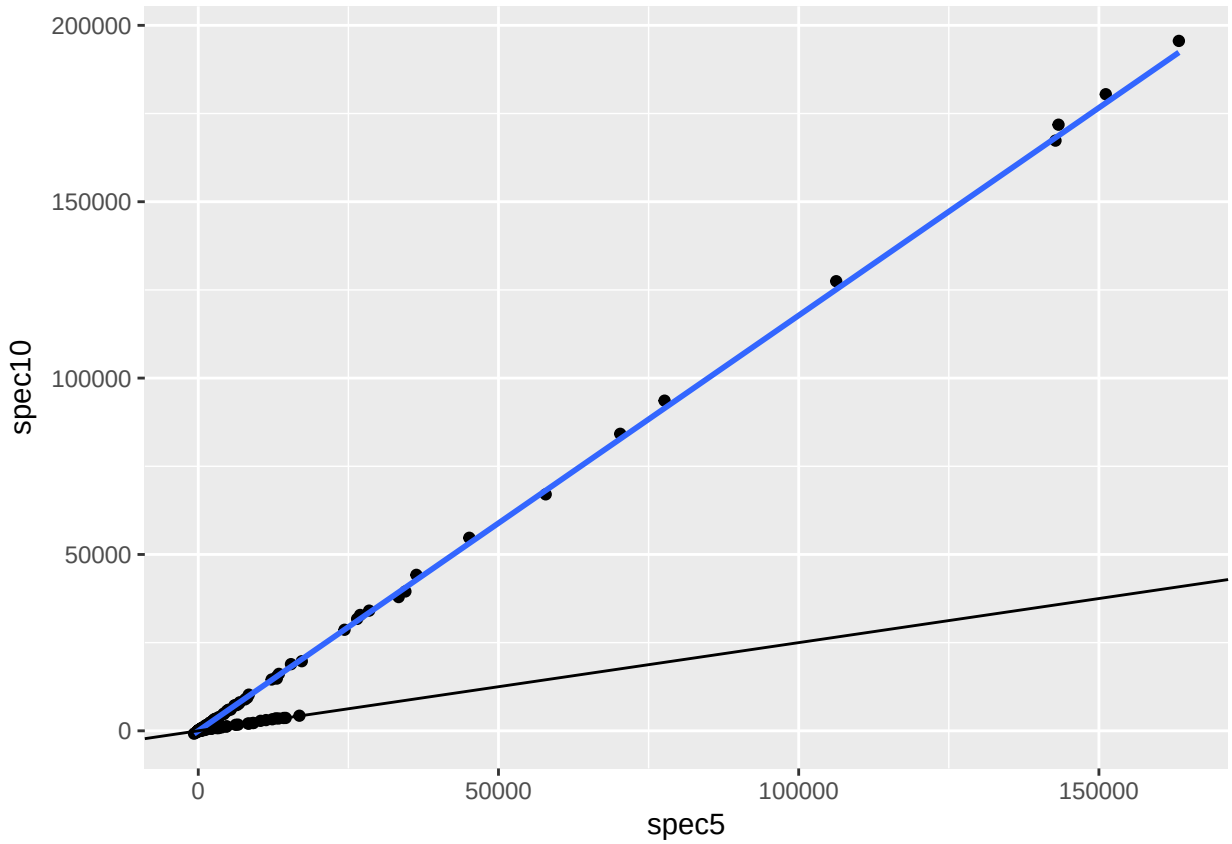
The second graph shows the frequency of measurements for both spec5 (red) and spec10(blue). They

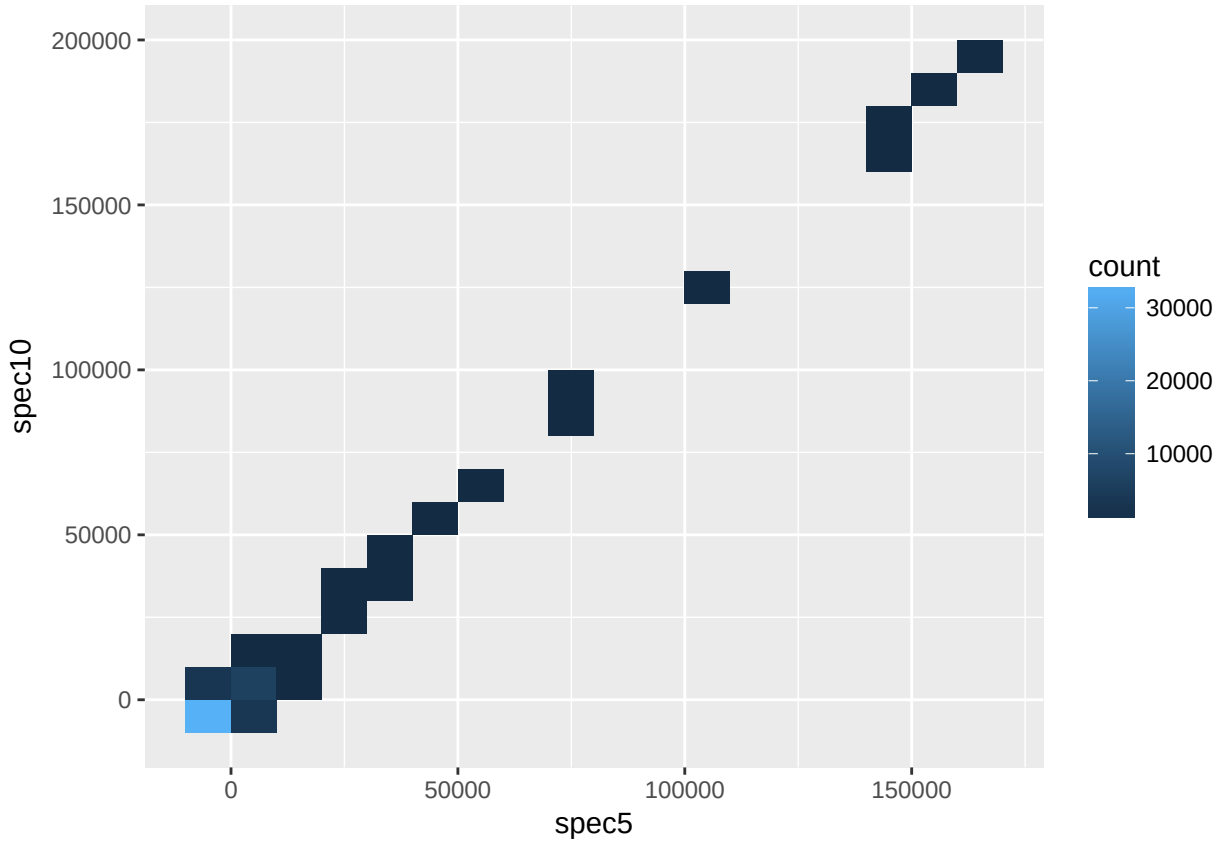
have similar trends, a higher concentration of low spectra values, but spec5 occurs slightly more often for low spectra values. The final graph depicts the range of spec10 values, limited by 30, and the number of occurrences of that value within the data set. Similar to the spec5, most of the measurements of spec10 were for lower values, around 5, the number of values greatly decreases and spec10 values above 30 are very minimal, and therefore were not included in the graph.

Zhenlong Li

```
## [1] 0.9952781
```

```
## Warning: Ignoring unknown parameters: intersect
```





The two different plots shows the correlation between spec5 and spec10. They are in a linear relationship but

there are two different regression line, so I plot bin2d plot which shows that the relation between them also depends on the number of data points. Overall, spec5 and spec10 are in strong correlation according to the calculated covariance.

Scott Baker

The first two plots show the variation on a logarithmic scale for the spec5 and spec10 variables.

```
var(ms$spec5)
```

```
## [1] 2830876
```

```
var(ms$spec10)
```

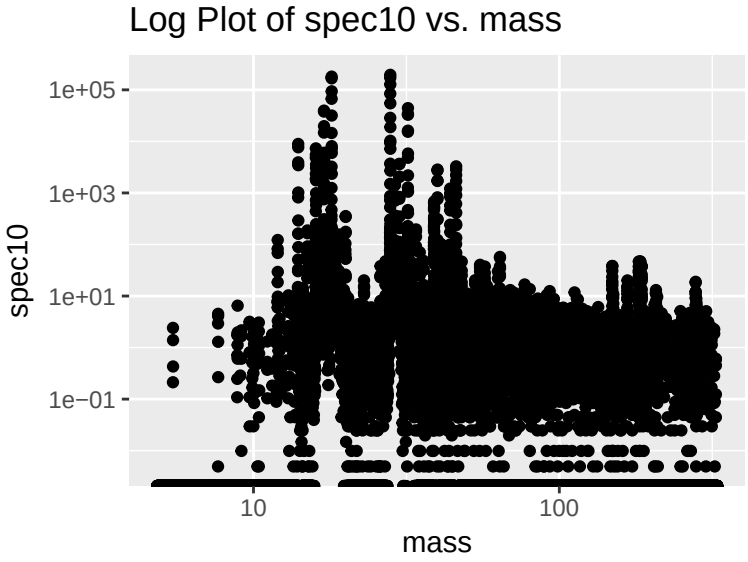
```
## [1] 3963210
```

Here is a plot of the relationship between the spec10 and mass variables on a logarithmic scale:

```
## Warning in self$trans$transform(x): NaNs produced
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
```

```
## Warning: Removed 838 rows containing missing values (geom_point).
```

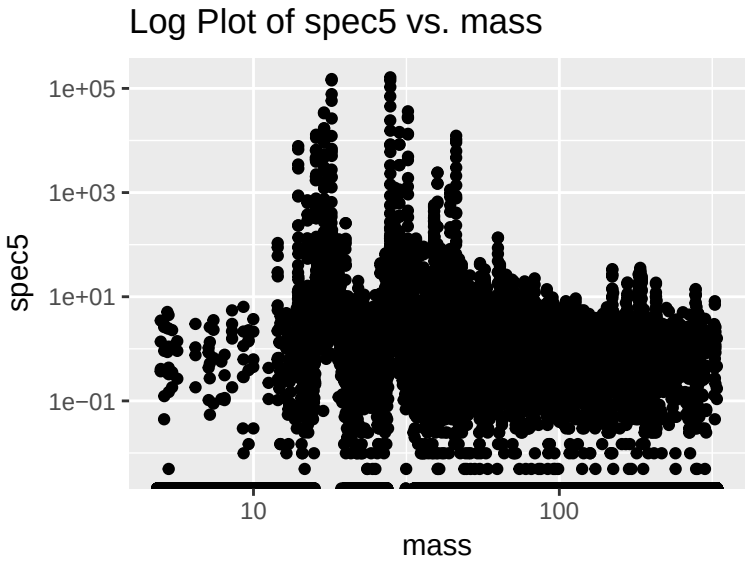


Here is a plot of the relationship between the spec5 and mass variables on a logarithmic scale:

```
## Warning in self$trans$transform(x): NaNs produced
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
```

```
## Warning: Removed 893 rows containing missing values (geom_point).
```



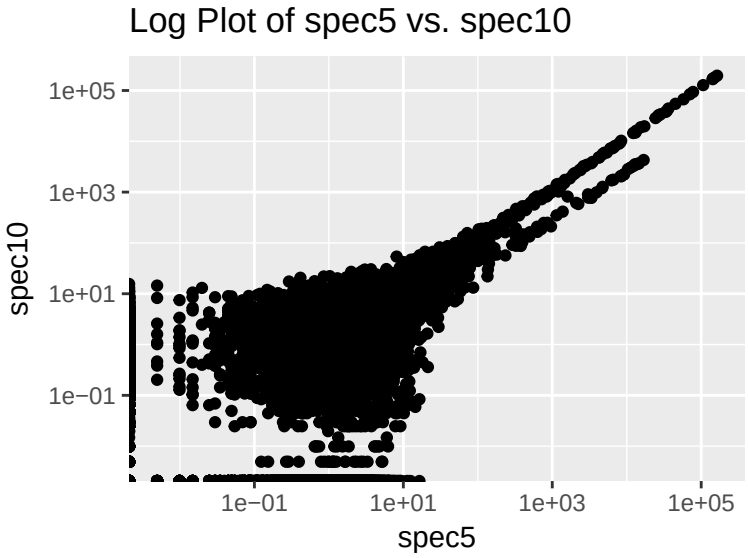
The covariance between the spec5 and spec10 is:

```
cor(ms$spec5,ms$spec10)
```

```
## [1] 0.9952781
```

Here is a plot of the relationship between the spec5 and spec10 variables on a logarithmic scale:

```
## Warning in self$trans$transform(x): NaNs produced
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning in self$trans$transform(x): NaNs produced
## Warning: Transformation introduced infinite values in continuous y-axis
## Warning: Removed 1690 rows containing missing values (geom_point).
```



It can be seen that there are highlights around 30 and fifty for mass. # Contributions * Lindsay: Created a histogram plot to see significant values of spec10, as well as to get a sense of the distribution of the data. For comparing spec5 to spec10, made a density plot to see how frequently each occured and to determine if one

spectra was more common than the other. To narrow the data filtered it all by positive values less than 50, it seemed to be the best range where majority of the data existed.

- Lexie: Created a histogram from raw spec5 data and a graph comparing the difference in spec5 and spec10 compared to spec5.
- Li: Created graphs about the correlation between spec5 and spec10 and improved the correlation with a bin2d plot.
- Scott: Created graphs on a logarithmic scale, since there is data on lots of orders of magnitude. Responsible for knitting this document and submitting to OSF. Finally, dealt with git organization and management.