Poster Presentation

By: Angus, Kevin, Yutong, Adriano

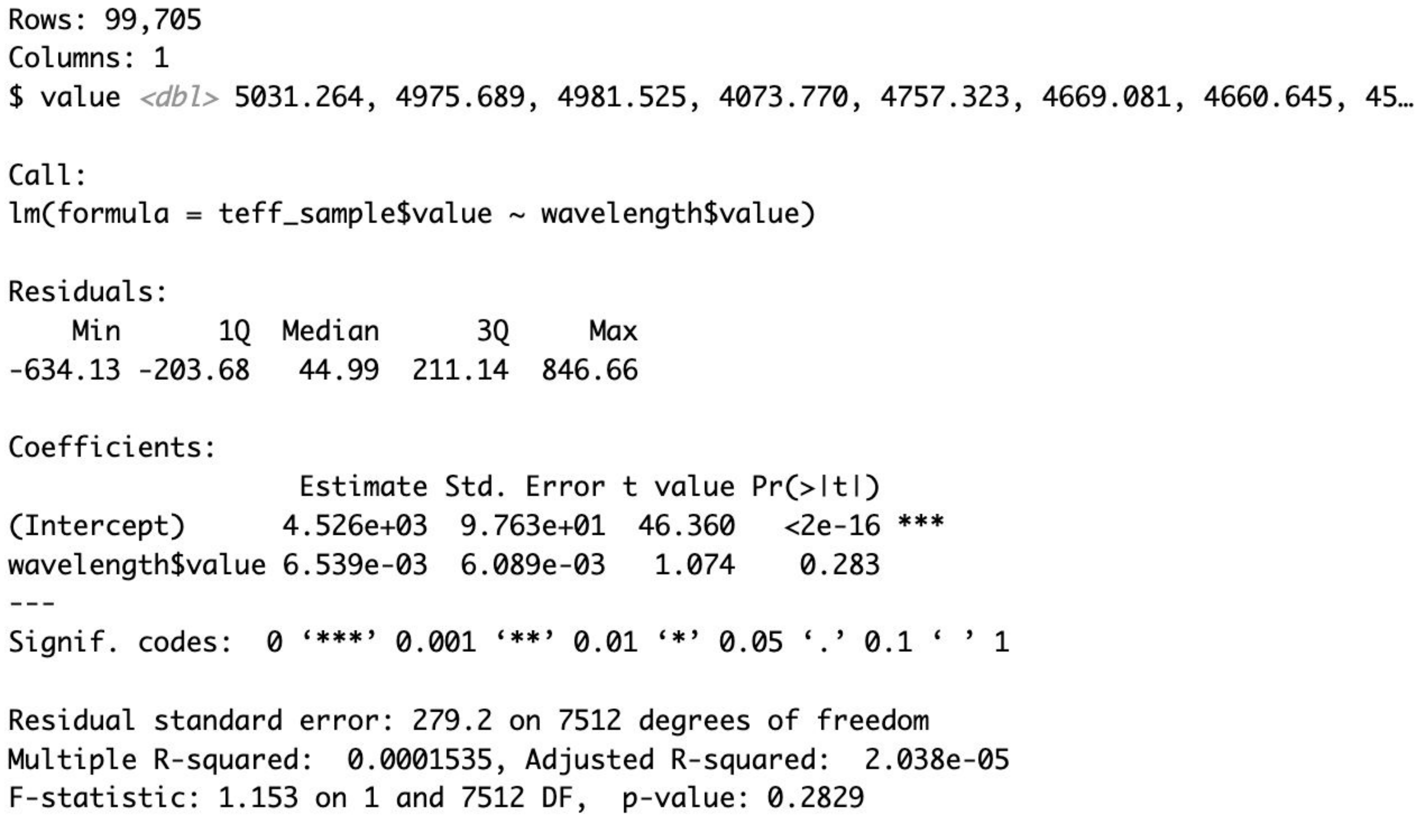
Abstract, Introduction, and Data

- We have chosen to analyze three questions based on the SDSS APOGEE Stellar Spectra data set
- Question 1: Does the effective temperature of the stars correlate to the wavelength?
  - Here, we are going to observe the effective temperatures and wavelengths of stars.
  - The wavelength data only contained 7514 observations, so we will take a random sample of effective temperature to match the wavelength count
  - For the experiment, we planned to perform a linear regression hypothesis test between the two values, with a standard p-value of 0.05
  - Important background information: Wien's Law (low effective temperatures of 3000K provided orange wavelengths, while at 30,000K provided blue wavelengths within stars).

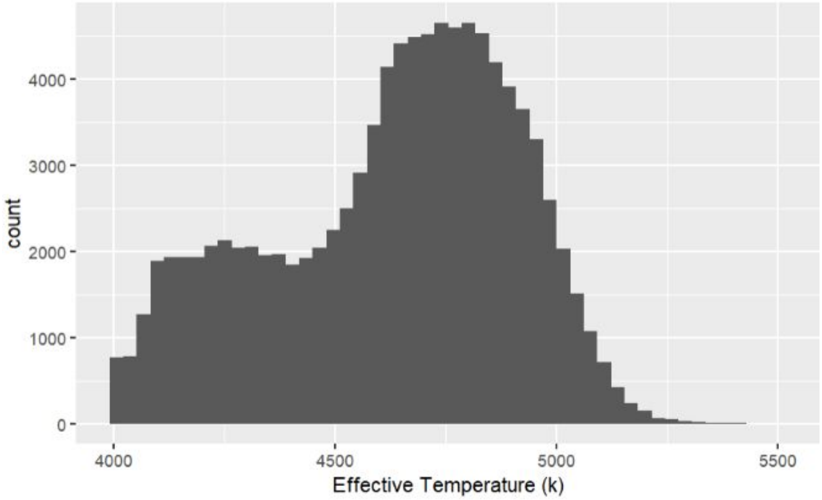
- Question 2: What is the estimation of the true population mean of the red giants' surface gravity?
  - In this question, we planned to observe the factors logg, which is the base 10 logarithm of the star, and is measured in centimeters-grams-seconds (cgs)
  - This is the only required data, since we only need to look at the surface gravity of red giant stars
  - As we discovered that iron is the most common element in stars
  - This data is computed through 1000 bootstrap samples, in order to compute confidence intervals for the true population mean of the red giant's surface gravity.
- Question 3: How does the range of the amount of iron, correlate to the amount of other elements that can be found on the surface of the stars?
  - For this question, we will compare the total iron mass in stars with the total mass of every other element in our data set
  - These include: iron, aluminum, carbon, calcium, iron, magnesium, nitrogen, and oxygen
  - We want to observe the correlation between iron and these elements and identify what this correlation represents.

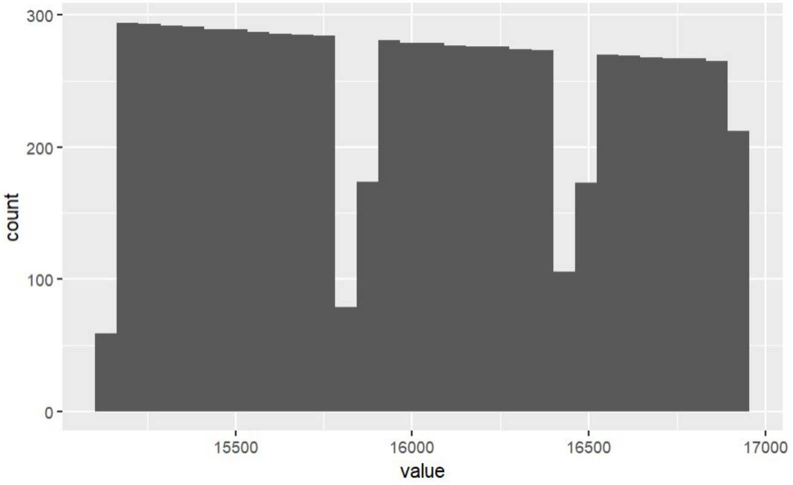
Question 1: Methods, Results & Visualizations

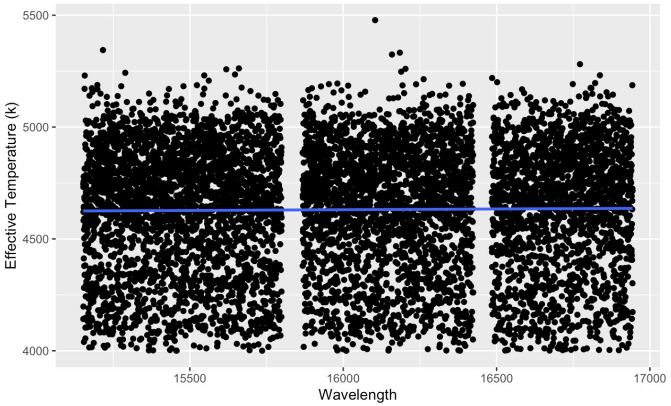
- Background Information: Wien’s Law
- Variables: Effective temperature (x), Wavelength (y)
- Clean data: na.omit()
- Hypothesis test of simple linear regression:
  - Null hypothesis: There is no correlation between wavelength and effective temperature for stars
  - Alternative hypothesis: There is a correlation between wavelength and effective temperature for stars
  - Same number of observations -> Use random sample of 7514 observations for teff, and entire wavelength population
- Statistical Data: lm() summary()
  - Residuals: 279.2 on 7512 degrees of freedom
  - P-value: 0.2829
- Since p-value > 0.05 and low residual errors, the data is not statistically significant, so we are not able to reject the null hypothesis. So, it is likely that there is no correlation between wavelength and effective temperature of stars.



```
Rows: 99,705
Columns: 1
$ value <dbl> 5031.264, 4975.689, 4981.525, 4073.770, 4757.323, 4669.081, 4660.645, 45…

Call:
lm(formula = teff_sample$value ~ wavelength$value)

Residuals:
    Min      1Q  Median      3Q     Max 
-634.13 -203.68   44.99  211.14  846.66 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      4.526e+03  9.763e+01  46.360   <2e-16 ***
wavelength$value 6.539e-03  6.089e-03   1.074    0.283    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 279.2 on 7512 degrees of freedom
Multiple R-squared:  0.0001535, Adjusted R-squared:  2.038e-05 
F-statistic: 1.153 on 1 and 7512 DF,  p-value: 0.2829
```

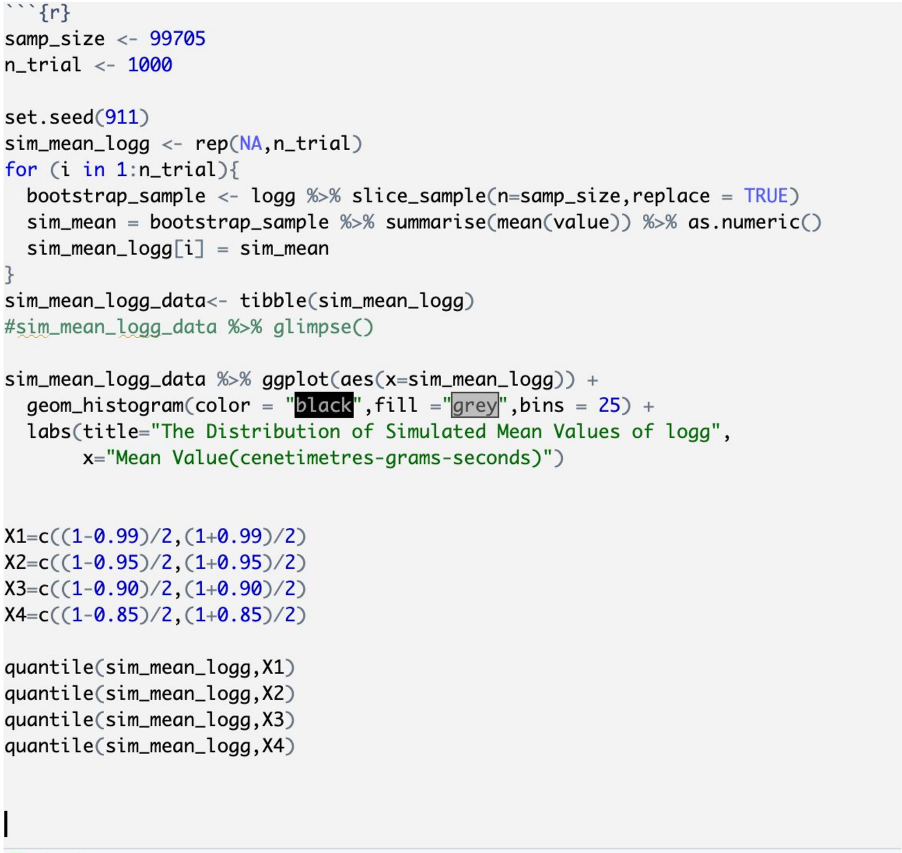

count
4000
3000
2000
1000
0
4000
4500
5000
5500
Effective Temperature (k)


count
300
200
100
0
15500
16000
16500
17000
value

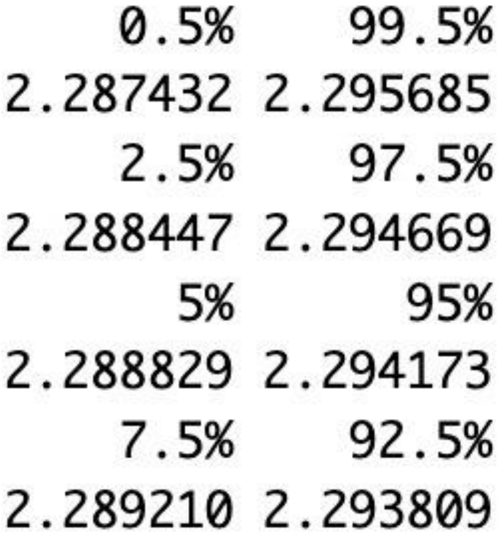

Effective Temperature (k)
5500
5000
4500
4000
15500
16000
16500
17000
Wavelength

Question 2: Methods, Results & Visualizations

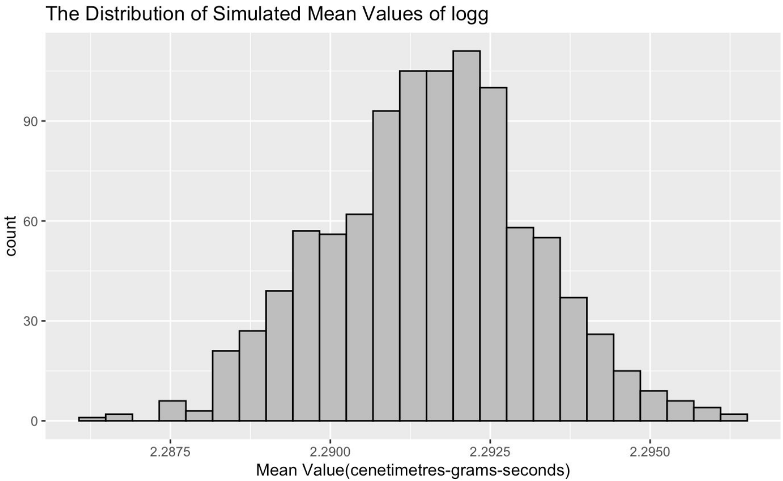
- Variable: logg (base-10 logarithm of surface gravity g of stars, where g is measured in cgs unit centimeters-grams-seconds.
- Clean data: !na() nrow()
- Simulation: bootstrap resampling, get the test statistic
- Visualization:
  - bootstrap sampling distribution histogram
  - bootstrap sampling distribution boxplot
- CI calculation
  - CI intervals of the 4 CI levels

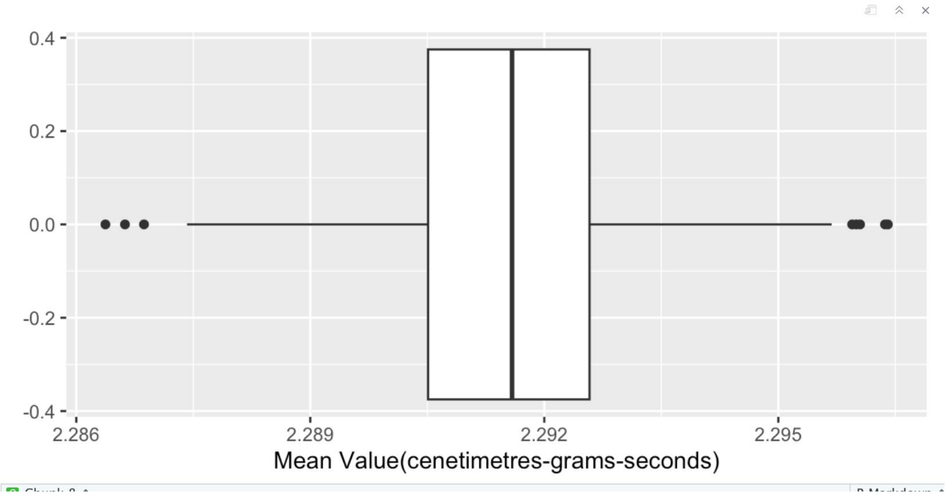


~~~
```{r}
samp_size <- 99705
n_trial <- 1000

set.seed(911)
sim_mean_logg <- rep(NA,n_trial)
for (i in 1:n_trial){
  bootstrap_sample <- logg %>% slice_sample(n=samp_size,replace = TRUE)
  sim_mean = bootstrap_sample %>% summarise(mean(value)) %>% as.numeric()
  sim_mean_logg[i] = sim_mean
}
sim_mean_logg_data<- tibble(sim_mean_logg)
#sim_mean_logg_data %>% glimpse()

sim_mean_logg_data %>% ggplot(aes(x=sim_mean_logg)) +
  geom_histogram(color = "black",fill ="grey",bins = 25) +
  labs(title="The Distribution of Simulated Mean Values of logg",
       x="Mean Value(cenetimetres-grams-seconds)")


X1=c((1-0.99)/2,(1+0.99)/2)
X2=c((1-0.95)/2,(1+0.95)/2)
X3=c((1-0.90)/2,(1+0.90)/2)
X4=c((1-0.85)/2,(1+0.85)/2)

quantile(sim_mean_logg,X1)
quantile(sim_mean_logg,X2)
quantile(sim_mean_logg,X3)
quantile(sim_mean_logg,X4)
~~~



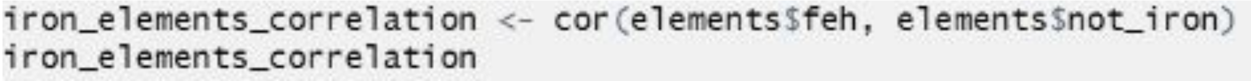
```
    0.5%    99.5% 
2.287432 2.295685 
    2.5%    97.5% 
2.288447 2.294669 
      5%      95% 
2.288829 2.294173 
    7.5%    92.5% 
2.289210 2.293809 
```


The Distribution of Simulated Mean Values of logg
count
0
30
60
90
2.2875
2.2900
2.2925
2.2950
Mean Value(cenetimetres-grams-seconds)

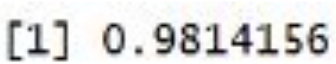

0.4
0.2
0.0
-0.2
-0.4
2.286
2.289
2.292
2.295
Mean Value(cenetimetres-grams-seconds)

Question 3: Methods, Results & Visualizations

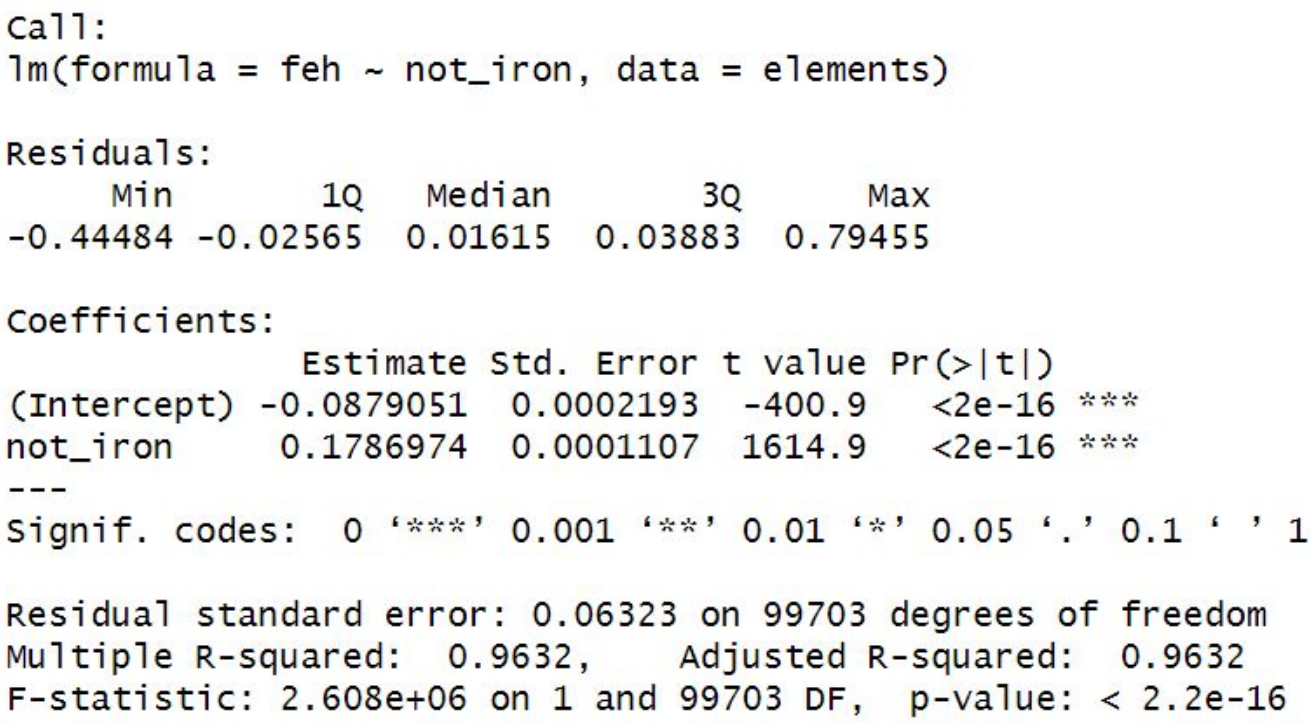
- We decided to take our iron to be one set of data, and combine the other elements in the data to be ‘not iron.
- Variables: x = abundance of iron, y = abundance of non-iron
- Correlation
    - cor() = 0.98
    - Close to 1
    - Strong positive correlation
- Visualization
    - Scatterplot
        - Line of best fit
    - Histogram
        - Unimodal
        - Left skew
- Statistical Data: lm() summary
    - Intercept = -0.08
    - Coefficient = 0.18



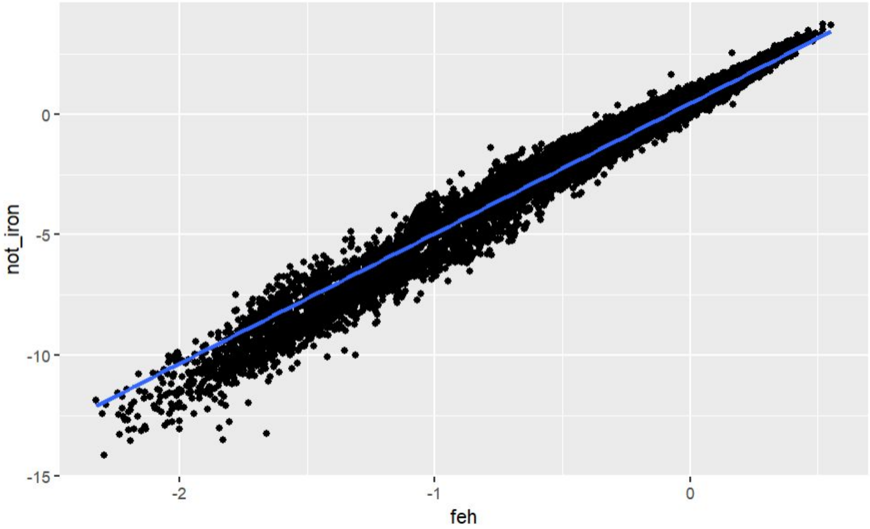
```
iron_elements_correlation <- cor(elements$feh, elements$not_iron)
iron_elements_correlation
```

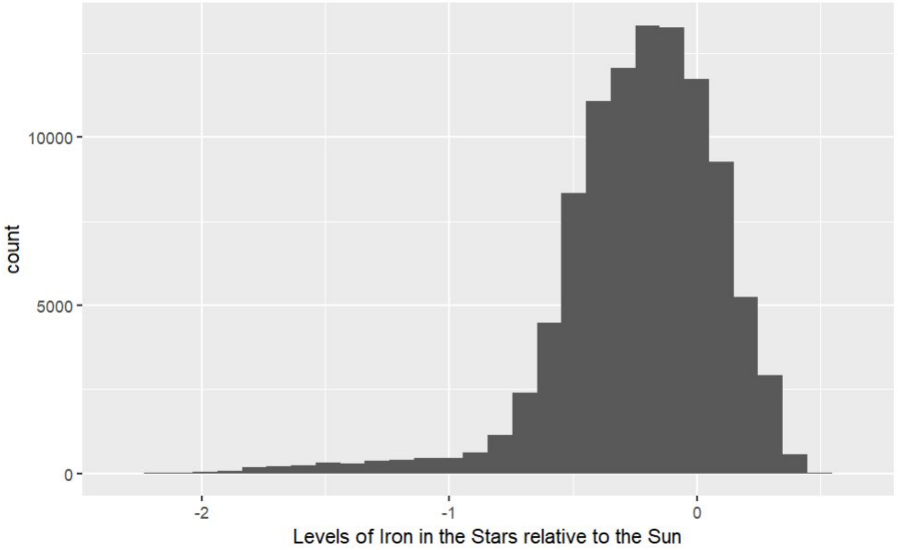


```
[1] 0.9814156
```



```
Call:
lm(formula = feh ~ not_iron, data = elements)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.44484 -0.02565  0.01615  0.03883  0.79455 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.0879051  0.0002193  -400.9   <2e-16 ***
not_iron     0.1786974  0.0001107  1614.9   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06323 on 99703 degrees of freedom
Multiple R-squared:  0.9632,	Adjusted R-squared:  0.9632 
F-statistic: 2.608e+06 on 1 and 99703 DF,  p-value: < 2.2e-16
```


0
-5
-10
-15
not_iron
-2
-1
0
feh


10000
5000
0
count
-2
-1
0
Levels of Iron in the Stars relative to the Sun

Conclusion and Discussion

- Question 1
  - Low Residuals and P-Value -> High Validity, Not Reject Null Hypothesis
  - Scatterplot: Horizontal Line of Best Fit
  - Other Methods: Bootstrapping, Residuals
- Question 2
  - Four different confidence intervals of 99%, 95%, 90%, and 85%
  - Our 99% confidence interval ranges from 2.288447cgs to 2.294669cgs
  - As our confidence interval decreases, the interquartile range also decreases, and therefore, our lower confidence intervals are subintervals of the 99% confidence interval
- Question 3
  - There is a strong positive correlation between iron and non-iron elements within the son
  - Scatterplot best line of fit represents this correlation
  - Results are statistically significant

Contributions

- Abstract, Intro, Data
  - Adriano Rymon-Lipinski
- Question 1
  - Kevin Shao
- Question 2
  - Yutong Han
- Question 3
  - Angus Yeung
- Conclusion
  - Adriano Rymon-Lipinski

Works Cited

1. Requirements of Capstone Project Given by STA130, The Dataset *SDSS APOGEE Stellar Spectra* As Given
2. https://www.atnf.csiro.au/outreach/education/senior/cosmicengine/stars_colour.html#:~:text=Wien%27s%20law%20is%20name%20of%20the%20direct%20relationship,and%20is%20a%20measure%20of%20the%20surface%20temperature. (Related to Wien’s Law, Used for Question 1’s Background Information)