Analyzing Factors Affecting Credit Limit

William Chettleburgh, Aaron Langtry, Charles Martel, and Jeremy Bouford

Introduction and Motivation

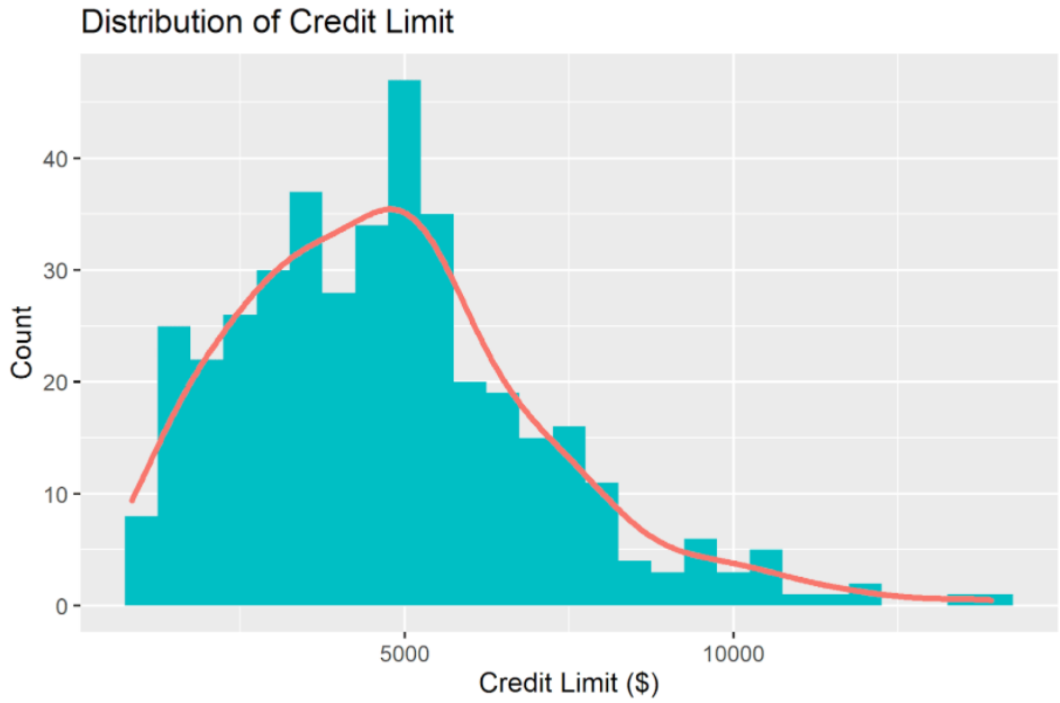
Using the Credit dataset provided by ISLR, we strove to answer:

- What financial and demographic factors contribute to having a higher or lower credit limit?
- Does gender or ethnicity play a role in having a higher credit limit?
- What are the most relevant and statistically significant variables for predicting credit limit?

```
Rows: 400
Columns: 12
$ ID        <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 3~
$ Income    <dbl> 14.891, 106.025, 104.593, 148.924, 55.882, 80.180, 20.996, 71.408, 15.125, 71.061, 63.095, 15.045, 80.616, 43.682, 19.144, 2~
$ Limit     <int> 3606, 6645, 7075, 9504, 4897, 8047, 3388, 7114, 3300, 6819, 8117, 1311, 5308, 6922, 3291, 2525, 3714, 4378, 6384, 6626, 2860~
$ Rating    <int> 283, 483, 514, 681, 357, 569, 259, 512, 266, 491, 589, 138, 394, 511, 269, 200, 286, 339, 448, 479, 235, 458, 213, 398, 156,~
$ Cards     <int> 2, 3, 4, 3, 2, 4, 2, 2, 5, 3, 4, 3, 1, 1, 2, 3, 3, 3, 1, 2, 4, 1, 3, 5, 3, 5, 6, 2, 2, 4, 4, 5, 2, 4, 2, 3, 2, 2, 4, 4, 3, 2~
$ Age       <int> 34, 82, 71, 36, 68, 77, 37, 87, 66, 41, 30, 64, 57, 49, 75, 57, 73, 69, 28, 44, 63, 72, 61, 48, 57, 25, 44, 44, 41, 55, 47, ~
$ Education <int> 11, 15, 11, 11, 16, 10, 12, 9, 13, 19, 14, 16, 7, 9, 13, 15, 17, 15, 9, 9, 16, 17, 10, 8, 15, 16, 12, 16, 14, 16, 5, 16, 13,~
$ Gender    <fct> Male, Female, Male, Female, Male, Male, Female, Male, Female, Female, Male, Male, Female, Male, Female, Female, Female, Fema~
$ Student   <fct> No, Yes, No, No, No, No, No, No, No, Yes, No, No, No, No, No, No, No, No, No, No, No, No, No, No, No, No, Yes, No, No, No, N~
$ Married   <fct> Yes, Yes, No, No, Yes, No, No, No, No, Yes, Yes, No, Yes, Yes, No, Yes, Yes, Yes, Yes, No, No, No, Yes, Yes, No, Yes, No, No~
$ Ethnicity <fct> Caucasian, Asian, Asian, Asian, Caucasian, Caucasian, African American, Asian, Caucasian, African American, Caucasian, Cauca~
$ Balance   <int> 333, 903, 580, 964, 331, 1151, 203, 872, 279, 1350, 1407, 0, 204, 1081, 148, 0, 0, 368, 891, 1048, 89, 968, 0, 411, 0, 671, ~
```

Credit Limit

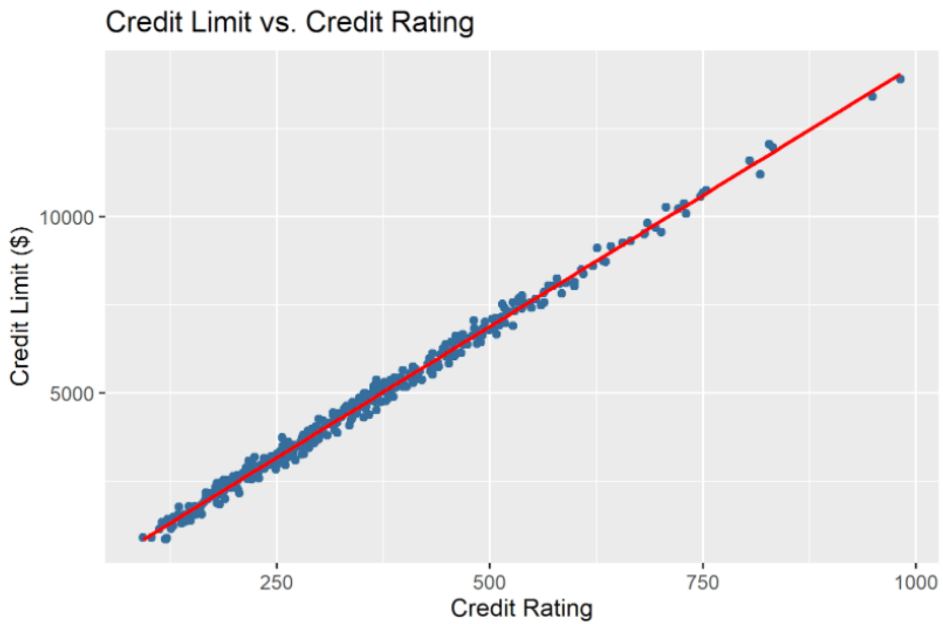
A *credit limit* is the maximum amount of "debt" a person can accumulate on a credit card before their purchases are denied. The credit limit varies from person to person, depending on how confident the company is that they will be paid back.



The distribution is **strongly skewed to the right**. As a result, when comparing two populations, a difference in medians test would be more applicable than a difference in means. However, many of the null distributions did not approach a normal curve with this test, so instead we use a non-parametric test called the Mann-Whitney U Test to test the hypotheses:

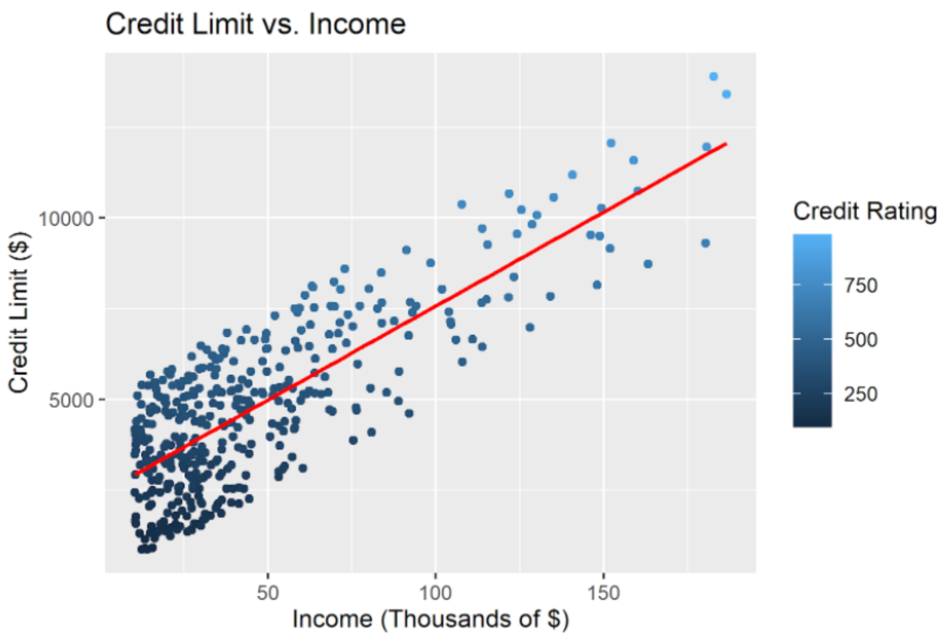
- H_0: The distributions of credit limit from which the two samples are drawn are the same (don't differ by a shift)
- H_A: The distributions of credit limit from which the two samples are drawn are different (do differ by a shift)

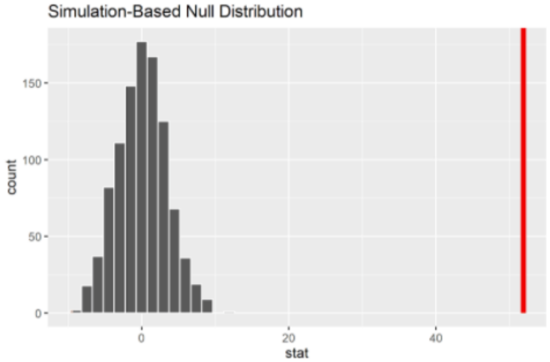
Impact of Credit Rating




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -542.92823   22.85026  -23.76   <2e-16 ***
Rating        14.87161    0.05903  251.95   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 182.4 on 398 degrees of freedom
Multiple R-squared:  0.9938,	Adjusted R-squared:  0.9938
F-statistic: 6.348e+04 on 1 and 398 DF,  p-value: < 2.2e-16
```

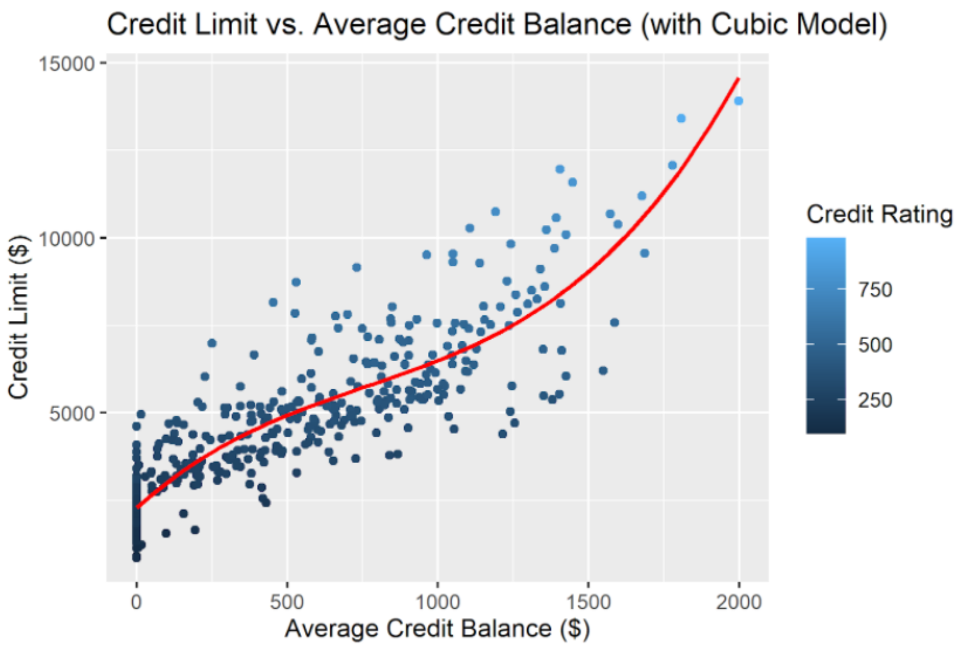
Impact of Income





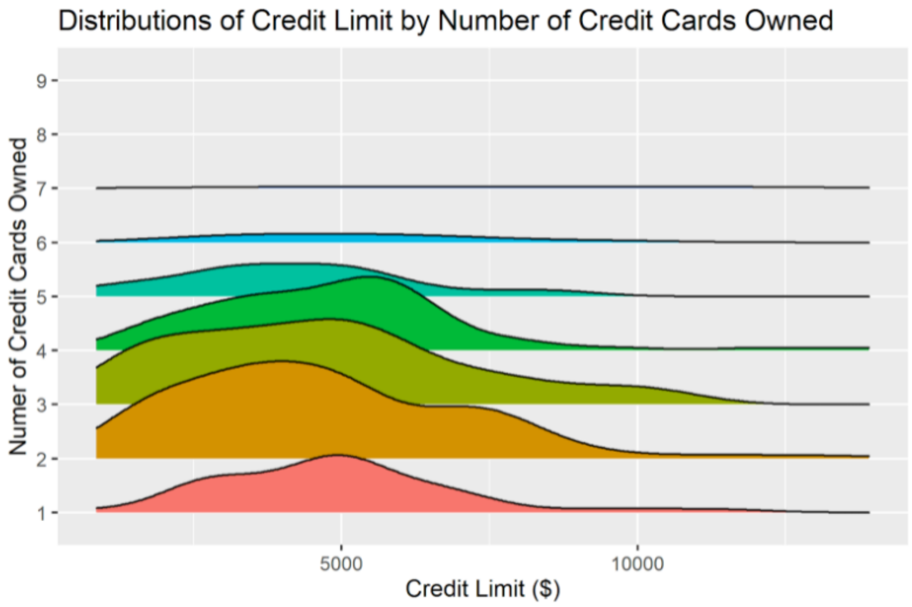

```
Coefficients:
            Estimate Std. Error
(Intercept) 2389.869    114.829
Income        51.875      2.004

p_value: 0
```

Impact of Credit Balance



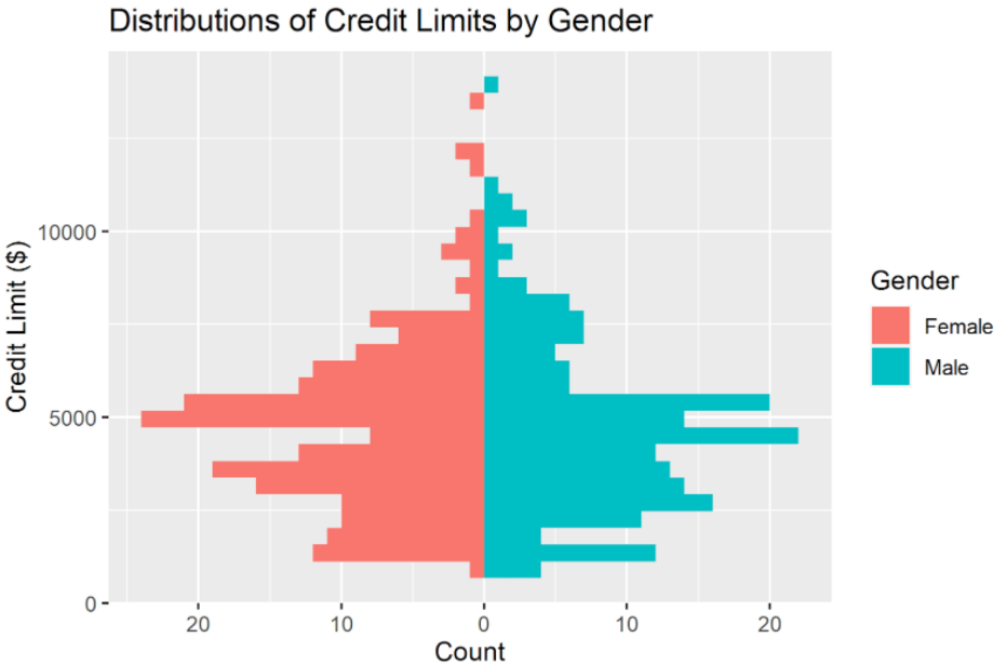

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              2.293e+03  1.065e+02  21.519  < 2e-16 ***
poly(Balance, 3, raw = T)1  7.708e+00  7.225e-01  10.669  < 2e-16 ***
poly(Balance, 3, raw = T)2 -6.235e-03  1.154e-03  -5.403 1.13e-07 ***
poly(Balance, 3, raw = T)3  2.729e-06  4.814e-07   5.668 2.79e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1130 on 396 degrees of freedom
Multiple R-squared:  0.762,	Adjusted R-squared:  0.7602
F-statistic: 422.7 on 3 and 396 DF,  p-value: < 2.2e-16
```

Impact of Number of Cards



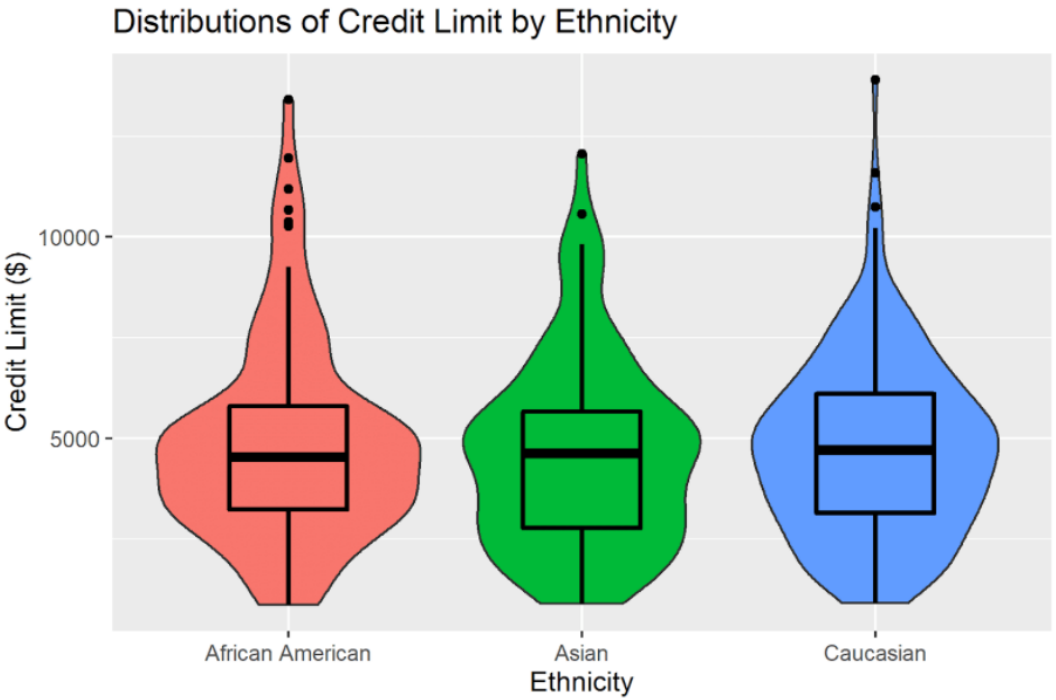

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4684.67     274.98  17.037   <2e-16 ***
Cards          17.22      84.37   0.204    0.838
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2311 on 398 degrees of freedom
Multiple R-squared:  0.0001047,	Adjusted R-squared:  -0.002408
F-statistic: 0.04167 on 1 and 398 DF,  p-value: 0.8384
```

Impact of Gender



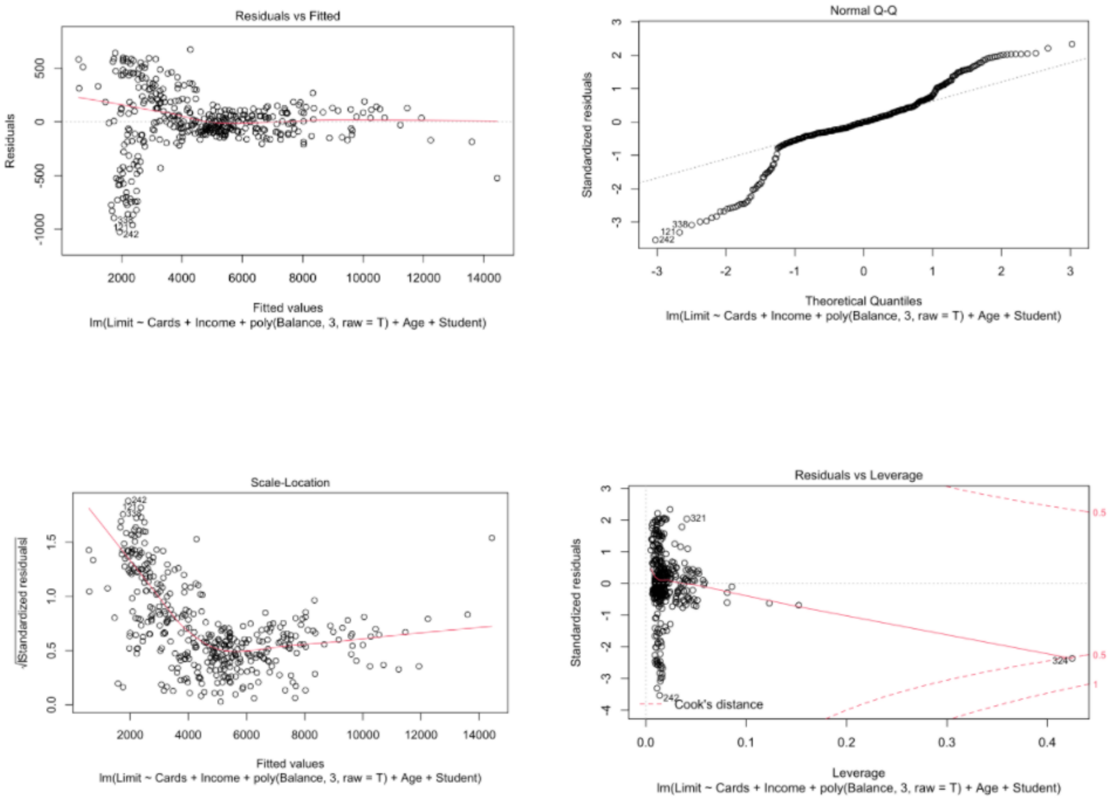

```
      Wilcoxon rank sum test with continuity correction

data:  Limit by Gender
W = 20532, p-value = 0.6304
alternative hypothesis: true location shift is not equal to 0
```

Impact of Ethnicity



	p value
Caucasian.v.Asian	0.4698917
Caucasian.v.African American	0.9954448
Asian.v.African American	0.5274932

Full Fitted Model (Without Credit Rating)

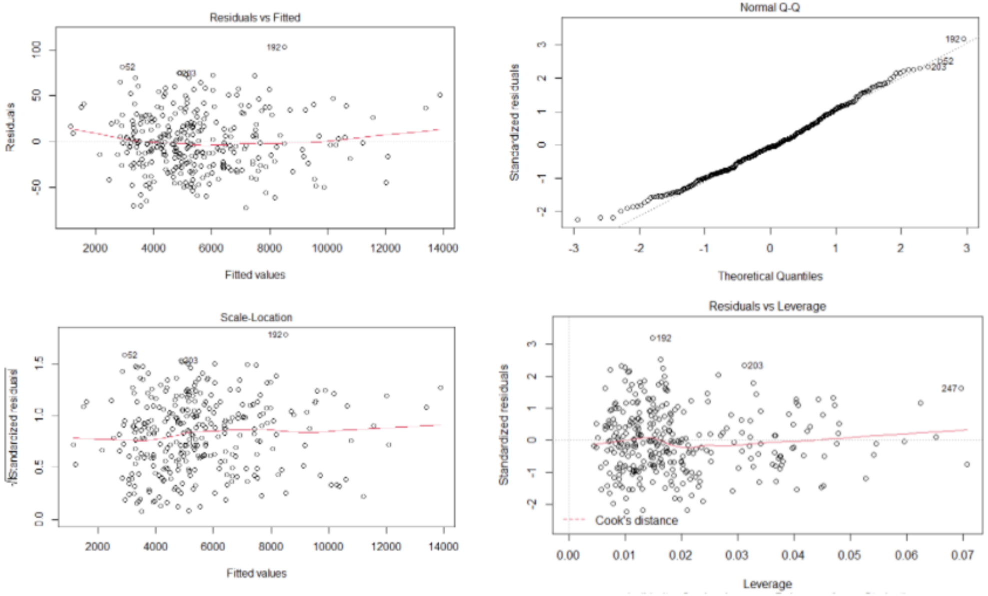



```
Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.510e+03  6.315e+01  23.919  < 2e-16 ***
Cards                         -7.330e+01  1.078e+01  -6.799 3.94e-11 ***
Income                         3.142e+01  5.058e-01  62.121  < 2e-16 ***
poly(Balance, 3, raw = T)1     5.905e+00  1.884e-01  31.339  < 2e-16 ***
poly(Balance, 3, raw = T)2    -3.336e-03  3.009e-04 -11.086  < 2e-16 ***
poly(Balance, 3, raw = T)3     1.101e-06  1.264e-07   8.710  < 2e-16 ***
Age                            2.128e+00  8.659e-01   2.458   0.0144 *
StudentYes                    -1.419e+03  5.124e+01 -27.688  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 291.7 on 392 degrees of freedom
Multiple R-squared:  0.9843,	Adjusted R-squared:  0.984
F-statistic:  3513 on 7 and 392 DF,  p-value: < 2.2e-16
```

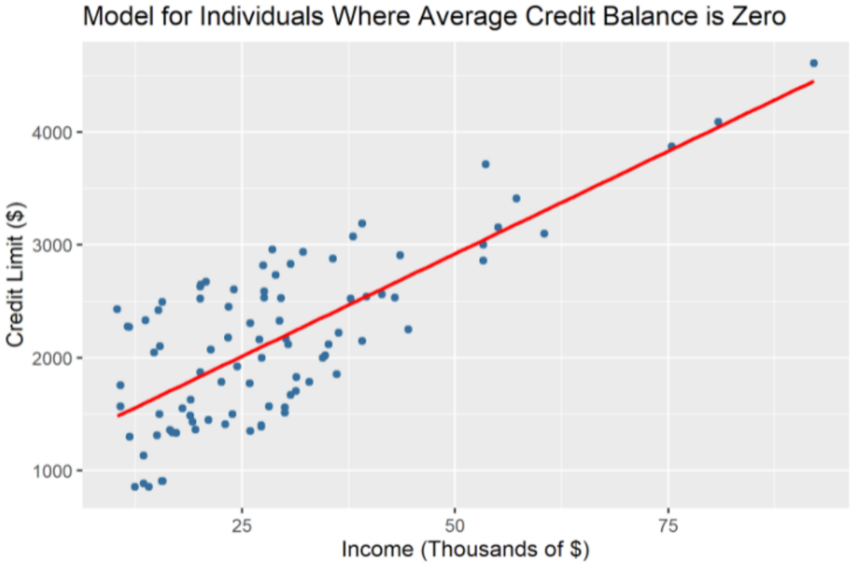
- The Normal Q-Q plot does not fall along a straight line. Since the tails twist counterclockwise, this suggests that the distribution has “heavy tails” (leptokurtosis)
- The leverage plot displays several points with high leverage, one of which has a Cook’s Distance less than -0.5 (and is thus is strongly influencing the model)
- The scale-location plot has significantly higher standardized residuals at lower fitted values of credit limit

Model for Balance > 0

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.151e+03  7.721e+00  278.60   <2e-16 ***
Cards       -7.593e+01  1.302e+00  -58.34   <2e-16 ***
Income       3.063e+01  5.573e-02  549.57   <2e-16 ***
Balance      3.061e+00  5.091e-03  601.17   <2e-16 ***
Age          3.067e+00  1.097e-01   27.97   <2e-16 ***
StudentYes  -1.533e+03  5.720e+00 -267.97   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.38 on 304 degrees of freedom
Multiple R-squared:  0.9998,     Adjusted R-squared:  0.9998
F-statistic: 2.483e+05 on 5 and 304 DF,  p-value: < 2.2e-16
```



Model for Balance == 0




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1104.691    111.647   9.895 5.92e-16 ***
Income        36.360      3.423  10.623  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 495.8 on 88 degrees of freedom
Multiple R-squared:  0.5619,     Adjusted R-squared:  0.5569
F-statistic: 112.8 on 1 and 88 DF,  p-value: < 2.2e-16
```

Only one variable (Income) was found significant for this model.

Conclusions

- Credit rating is an extremely strong predictor of credit balance.
- The credit limits for those with zero average credit balance are fundamentally different than those with a positive average credit balance.
- For those with a positive average credit balance (on average, all else held constant):
    - Expected decrease in credit limit by $76 per addition credit card
    - Expected increase in credit limit by $31 per additional $1000 in income
    - Expected increase in credit limit by $3 per additional dollar of average credit balance
    - Expected increase in credit limit by $3 per additional year in age
    - Expected decrease in credit limit by $1533 for students compared to non-students
- For those with no credit balance, income is the only significant predictor.
    - Expected increase in credit limit by $36 per additional $1000 in income
- No significant difference between credit limits was found based on gender and ethnicity -- age was the only demographic predictor which was significant.
    - The coefficient is still quite small, so it isn't particularly significant in practice

References

Climbs_lika_Spyder. (2016, October 21). *Emulate GGPLOT2 Default Color Palette*. Stack Overflow. Retrieved December 5, 2021, from https://stackoverflow.com/a/40181166.

gung - Reinstate Monica. (2013, March 14). *QQ Plot Does Not Match Histogram*. Cross Validated. Retrieved December 5, 2021, from https://stats.stackexchange.com/a/52221.

Kim, B. (2015, September 21). *Understanding Diagnostic Plots for Linear Regression Analysis*. Research Data Services + Sciences. Retrieved December 5, 2021, from https://data.library.virginia.edu/diagnostic-plots/.

strictlystat. (2014, June 10). *Making Back-to-Back Histograms*. A HopStat and Jump Away. Retrieved December 5, 2021, from https://hopstat.wordpress.com/2014/06/10/making-back-to-back-histograms/.