# Special Topics in Data Science Using R – Student Feedback

The most surprising part of this course was the detailed data munging (data cleaning) that had to be done to get the data ready for use in R. [0.5984] Great emphasis was placed on not removing data unless there was ample reason to consider it tainted (i.e., not merely because an attention check was missed or the data point was an outlier). [0.4404] Additionally, extreme care and impartiality were used when translating open-response (qualitative) data into numerical values R could interpret. [0.7096]

Finally, the opportunity to apply exploratory data analysis to real-world data was such a great learning experience. [0.7845] Testing variables for normality and comparing data sets with and without observations that missed the attention or effort checks gave me increased familiarity with R coding and graphical interpretation, besides providing valuable insights about the effects of certain flagged observations. [0.8013] Our team also examined outliers, demographics, and correlations. [0.0] Without a doubt, this course has been a valuable and memorable learning experience. [.6381]

- Faith V., Student at Ouachita Baptist University, 2021

|  |  |  |  |
| --- | --- | --- | --- |
| **Text Analysis** | | | |
|  |  | Sentiment (Opinion Mining)1 | Tone Mix Classification (Emotion Mining)2 |
|  | Para. 1 | 0.5828 (Pos) | Happy 25%, Angry 0%, Surprise 25%, Sad 12%, Fear 38% |
|  | Para. 2 | 0.5559 (Pos) | Happy 43%, Angry 0%, Surprise 14%, Sad 14%, Fear 29% |
|  | Para. 1 & 2 Combined | 0.5675 (Pos) | Happy 33%, Angry 0%, Surprise 20%, Sad 13%, Fear 33% |

Notes: 1Processed in Python via VADER. 2Processed via Web Analytic Application (<https://text2emotion.herokuapp.com>).

Over the summer, I was able to get hands-on experience in research and data analytics while studying the integration of Slack into higher education. [0.0] I was able to apply my introductory knowledge of R to help clean data sets, organize groups of variables, and prepare plots and graphs. [.6597] The biggest thing that I learned throughout the course of this summer was the importance of having sound data in the beginning. [.3612] Research will eventually fall through and not hold up to statistical testing if it isn’t collected properly. [0.0]

I also learned that even though it is extremely time-consuming, cleaning the data properly is an essential step in the research process. [0.0] Without clean, organized data, you can’t even begin to perform tests or code graphs. [-.3089] In addition to learning how to properly clean data, I also learned how crucial graphs and plots can be. [.4019] A graph is able to take hundreds of seemingly meaningless data points and put them together and tell a story or demonstrate a relationship. [-.4404] R is an amazing coding language for graphing because of the ggplot2 software in the tidyverse package. [.5859] I was able to see firsthand how to create complex QQ plots that integrate different variables, and these graphs helped piece the data together. [-.4278] Ultimately, this summer showed me that the process of research starts with data, deals with lots of data and ends with trying to demonstrate how that data is correlated. [0.0]

- Jeff M., Student at Ouachita Baptist University, 2021

|  |  |  |  |
| --- | --- | --- | --- |
| **Text Analysis** | | | |
|  |  | Sentiment (Opinion Mining)1 | Tone Mix Classification (Emotion Mining)2 |
|  | Para. 1 | 0.2552 (Pos) | Happy 43%, Angry 0%, Surprise 57%, Sad 0%, Fear 0% |
|  | Para. 2 | -0.0200 (Neg-Neu) | Happy 19%, Angry 0%, Surprise 56%, Sad 12%, Fear 12% |
|  | Para. 1 & 2 Combined | 0.0756 (Neu) | Happy 26%, Angry 0%, Surprise 57%, Sad 9%, Fear 9% |

Notes: 1Processed in Python via VADER. 2Processed via Web Analytic Application (<https://text2emotion.herokuapp.com>).