**DSC 520 READING ASSIGNMENT LIST**

1. **Introduction to R**

References:

This book is for those of you who are used to Python but maybe not so experienced with R. You will find a lot of help with Greg Wilson’s work in general and I highly recommend seeking his other work.

* Wilson, Greg. (2019). [The TidyNomicon](https://gvwilson.github.io/tidynomicon/).

Next, this course is all about statistics, specifically inferential or frequentist statistics. These concepts also form the basis of much of machine learning and every other buzzword that we have about what computers can do. Your textbook will provide you with a lot of background about stats but you will always need more knowledge. Each week, I will provide a “Refresher” of sorts from around Data Science. This is the first one:

* For those who need a refresher:
  + Vallantin, Wilame Lima. (2018). [Statics and Probability Refresher](https://towardsdatascience.com/statistics-and-probability-refresher-fb642a3a5f7d).
* For those who need a lot of refreshing:
  + Kunin, Daniel. [Seeing Theory](https://seeing-theory.brown.edu/).
* If you need additional help with importing data, check out this brief tutorial from rStudio itself:
  + Luraschi, Javier. (2019). [Importing Data with RStudio](https://support.rstudio.com/hc/en-us/articles/218611977-Importing-Data-with-RStudio).

1. **Summarizing Data; Setting up Project and Using Scripts**

If you need additional help, check out this brief tutorial from rStudio itself:

[Importing Data with RStudio](https://support.rstudio.com/hc/en-us/articles/218611977-Importing-Data-with-RStudio). RStudio

1. **Data Plotting:**

You can find additional help with ggplot2 in the following resources:

* Wickham, Hadley. (2014). [Ggplot2 2.2.0](https://blog.rstudio.com/2016/11/14/ggplot2-2-2-0/).
* Prabhakaran, Selva. [Ggplot2 – Essentials](http://www.sthda.com/english/wiki/ggplot2-essentials).
* STHDA. [Ggplot2 – Essentials](http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html).

There are a ton of other resources out there!

1. **Transformations and RMarkdown**

* Supplemental Reading
  + [R Markdown from R Studio](https://rmarkdown.rstudio.com/lesson-1.html), RStudio
  + [R Markdown Cheat Sheet from RStudio](http://content.bellevue.edu/cst/dsc/520/id/resources/rmarkdown-cheatsheet-2.0.pdf), RStudio
* Week 4 Statistics Refresher and Hints

This book will provide you with a lot of help if you give it some time:

* + Xie, Yihui; Allaire, J. J.; & Grolemund, Garrett. (2019). [R Markdown: The Definitive Guide](https://bookdown.org/yihui/rmarkdown/).

Also, don’t be afraid to poach things from:

* + RStudio. [RStudion. R Markdown](https://rmarkdown.rstudio.com/lesson-1.html" \t "_blank).

1. **Correlation**

Week 5 Readings, Assignments, and Tasks

Here is a breakdown of your assignments and tasks this week:

* 1. Read Chapter 6 from *Discovering Statistics Using R* and Chapter 18, Sections 18.1 from *R for Everyone*.
  2. DataCamp Exercise 5: Visualizing Two Variables; Correlation
  3. 5.1 Assignment: Student Survey
  4. 5.2 Discussion: Tech Support Fort

Supplemental Reading

* 1. [R Markdown from RStudio](https://rmarkdown.rstudio.com/lesson-1.html), RStudio
  2. [R Markdown Cheat Sheet from RStudio](http://content.bellevue.edu/cst/dsc/520/id/resources/rmarkdown-cheatsheet-2.0.pdf), RStudio

Week 5 Statistics Refresher and Hints

For those of you wondering about correlation, there are a lot of examples and ideas around the web. I find this one particularly useful:

* 1. Magnusson, Kristoffer. [Interpreting Correlations](https://rpsychologist.com/d3/correlation/).

Similarly, this one is also useful:

* 1. Wagih, Omar. [Guess the Correlation](http://guessthecorrelation.com/).

Finally, this item from week 1 has some discussion about correlation:

* 1. Kunin, Daniel. [Seeing Theory](https://seeing-theory.brown.edu/).

1. **Getting Started with Regression**

* Week 6 Readings, Assignments, and Tasks

Here is a breakdown of your assignments and tasks this week:

* + Read Chapter 7: Sections 7.1-7.5 from *Discovering Statistics Using R* and Chapter 18: Section 18.2 from *R for Everyone*.
  + DataCamp Exercise 6: Simple Linear Regression, Interpreting Regression Models, Model Fit
  + 6.1 Assignment: GSS 2016 Survey Data
  + 6.2 Discussion: Tech Support Fort
* Supplemental Reading
  + [R Markdown from R Studio](https://rmarkdown.rstudio.com/lesson-1.html), RStudio
* Week 6 Statistics Refresher and Hints

Generally, regression takes a lot of effort to get right and very little effort to get wrong. It is the thing you’ll study the most yet it is the thing you’ll remember the least. Here are some readings about regression for those of you who haven’t had to deal with it before:

* + Gallo, Amy. (2015). [A Refresher on Regression Analysis](https://hbr.org/2015/11/a-refresher-on-regression-analysis).
  + Martin, Rose. (2018). [Using Linear Regression for Predictive Modeling in R](https://www.dataquest.io/blog/statistical-learning-for-predictive-modeling-r/).
  + Statistics How To. [Regression Analysis: Step by Step Articles, Videos, Simples Definitions](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/regression-analysis/).
* DataCamp Exercise 6: Simple Linear Regression; Interpreting Regression Models; and Model Fit
  + Complete [Simple Linear Regression](https://www.datacamp.com/groups/dsc-520/assignments/23087) in DataCamp.
  + Complete [Model Fit](https://www.datacamp.com/groups/dsc-520/assignments/23089) in DataCamp.
  + Complete [Interpreting Regression Models](https://www.datacamp.com/groups/dsc-520/assignments/23088) in DataCamp.

1. **Multiple Regression**

Week 7 Readings, Assignments, and Tasks

Here is a breakdown of your assignments and tasks this week:

* 1. Read Chapter 7: Sections 7.6-7.12.2 from *Discovering Statistics Using R* and Chapters 19: Section 19.2-19.3 from *R for Everyone*.
  2. 7.1 Assignment: Housing Data
  3. 7.2 Discussion: Collaborate and Ask Questions

Week 7 Statistics Refresher and Hints

Generally, regression takes a lot of effort to get right and very little effort to get wrong. It is the thing you’ll study the most yet it is the thing you’ll remember the least. Here are some readings about regression for those of you who haven’t had to deal with it before:

* 1. Gallo, Amy. (2015). [A Refresher on Regression Analysis](https://hbr.org/2015/11/a-refresher-on-regression-analysis).
  2. Martin, Rose. (2018). [Using Linear Regression for Predictive Modeling in R](https://www.dataquest.io/blog/statistical-learning-for-predictive-modeling-r/).
  3. Statistics How To. [Regression Analysis: Step by Step Articles, Videos, Simples Definitions](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/regression-analysis/).

DataCamp Exercise 7: Multiple & Logistic Regression

Complete [Multiple and Logistic Regression](https://www.datacamp.com/courses/multiple-and-logistic-regression) in DataCamp.

1. **Logistic Regression**

* Week 8: Readings, Assignments, and Tasks

Here is a breakdown of your assignments and tasks this week:

* + Read:
    - Chapter 8 from *Discovering Statistics Using R.*
    - [Understanding the Bias-Variance Tradeoff](http://scott.fortmann-roe.com/docs/BiasVariance.html), Scott Fortmann-Roe, 2012
    - [Calculating UAC: The Area Under a ROC Curve](https://www.r-bloggers.com/calculating-auc-the-area-under-a-roc-curve/), Bob Horton, R-bloggers, 2016
  + 8.1 Assignment: Logistic Regression
  + 8.2 Assignment: Fit Logistic Regression Model
  + 8.3 Discussion: Tech Support Fort
* Week 8 Statistics Refresher and Hints

Again, you may find these resources useful this week:

* + Gallo, Amy. (2015). [A Refresher on Regression Analysis](https://hbr.org/2015/11/a-refresher-on-regression-analysis).
  + Martin, Rose. (2018). [Using Linear Regression for Predictive Modeling in R](https://www.dataquest.io/blog/statistical-learning-for-predictive-modeling-r/).
  + Statistics How To. [Regression Analysis: Step by Step Articles, Videos, Simples Definitions](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/regression-analysis/).

And also, this:

* + UCLA Statistical Consulting Group. [R Data Analysis Examples](https://stats.idre.ucla.edu/r/dae/logit-regression/).

This video is also great. I watch it when I need to remember what the heck logistic regression even is:

* + Simplilearn. (2018). Logistic Regression in R.

Generally, regression takes a lot of effort to get right and very little effort to get wrong. It is the thing you’ll study the most yet it is the thing you’ll remember the least. Here are some readings about regression for those of you who haven’t had to deal with it before:

* Gallo, Amy. (2015). [A Refresher on Regression Analysis](https://hbr.org/2015/11/a-refresher-on-regression-analysis).
* Martin, Rose. (2018). [Using Linear Regression for Predictive Modeling in R](https://www.dataquest.io/blog/statistical-learning-for-predictive-modeling-r/).
* Statistics How To. [Regression Analysis: Step by Step Articles, Videos, Simples Definitions](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/regression-analysis/).

1. **Machine Learning**

Week 9 Readings, Assignments, and Tasks

Here is a breakdown of your assignments and tasks this week:

* Read the following:
  + Machine Learning Fundamentals
    - Bernard Marr. (2016). [Supervised V Unsupervised Machine Learning – What’s The Difference?](https://www.forbes.com/sites/bernardmarr/2017/03/16/supervised-v-unsupervised-machine-learning-whats-the-difference/#4d2a4cf2485d)
    - Bernard Marr. (2016). [What Is The Difference Between Artificial Intelligence And Machine Learning?](https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/)
    - Bernard Marr. (2016). [What Is The Difference Between Deep Learning, Machine Learning and AI?](https://www.forbes.com/sites/bernardmarr/2016/12/08/what-is-the-difference-between-deep-learning-machine-learning-and-ai/)
  + K-Means Clustering
    - Sejal Jaiswal. (2018). [K-Means Clustering in R Tutorial](https://www.datacamp.com/community/tutorials/k-means-clustering-r)
    - Andrea Trevino. [Introduction to K-means Clustering](https://www.datascience.com/blog/k-means-clustering)
  + Nearest Neighbors Classification
    - Kevin Zakka. (2016). [A Complete Guide to K-Nearest-Neighbors with Applications in Python and R](https://kevinzakka.github.io/2016/07/13/k-nearest-neighbor/)
    - Scikit Learn. [Nearest Neighbors Classification](http://scikit-learn.org/stable/modules/neighbors.html#classification)
* 9.1 Discussion: Final Project Step 1: Getting Started
* 9.2 Assignment: Introduction to Machine Learning
* 9.3 Assignment: Clustering