**CMPSC 301**
**Data Analytics**
**Fall 2018**

**Lab 6: Statistical Analysis of Psychological Data**
**Using R and the R *Psych* package**
**2$^{nd}$ November 2018**

# Objectives

To explore statistical tools which are relevant for the evaluation of psychological data. In particular, to be able to research how to use new R-statistics software packages and apply them to particular contexts for which they were designed. To extract knowledge from the produced visualizations and extracted interpretation of results.

# Reading Assignment

Please review the technical manual for the *Psych* package for use with R. It can be found at the following link: `http://personality-project.org/r/overview.pdf`. This document will be used to guide your coding for testing and visualization.

# Groupwork

You are to work in a group of not more than four (4) people for this lab. Be sure to discuss each of the tasks and proceed after the group has come to a complete agreement. **Each person is to turn in his or her own report and code, however all lab partners should be listed in the submission.**

### GitHub Starter Link: Group work

<p align="center">https://classroom.github.com/g/TxBQkIPX</p>

To use this link, please follow the steps below.

- Your group leader will click on the link and accept the assignment and prepare a team name. All other members will later click on the link and select their team's name from the list that will appear.

- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab.

- Clone this repository (bearing your name) and work on the lab locally.

- As you are working on your lab, you are to commit and push regularly. You can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):

&ndash; `git commit <nameOfFile> -m ''Your notes about commit here''`

&ndash; `git push`

Alternatively, you can use the following commands to add multiple files from your repository:

&ndash; `git add -A`

&ndash; `git commit -m ''Your notes about commit here''`

&ndash; `git push`

# Psychology Data Analysis

In this lab you will utilize the *Psych* package to analyze data from the psychology discipline. In particular your deliverables will be composed of the following items.

## Obtaining Data

- **Obtain Your Data Set**: You may obtain your data from any online source as long as it is a credible source and that the data stems from the psychology discipline. For an idea, you could select one of the data sets available on the *Openpsychometrics* website: `https://openpsychometrics.org/_rawdata/` concerning personality types (Big Five Personality Test), Anxiety (Taylor Manifest Anxiety Scale), as well as other types of data stemming from psychological inquiry.

- **Describe Your Data**: You are to write a short report to describe your data. In your own words, discuss what the data is about: what it contains (what is measures) and why it is was collected. Be sure to locate the data's *code-book* to help you learn what the columns of data contain and to provide you with extra data about the set in general such as how the data was collected, from whom, and similar issues. In psychology, data sets are likely to have an associated article in which the data was featured. Look for articles for this extra information and use them if they exist.

## Method Research

Often in data analytics work, you will have to read library documentation to determine how to use particular methods for your work. In this section, you will spend some time studying the *Psych* library documentation (link: `http://personality-project.org/r/overview.pdf`) to learn how to run types of tests.

1. **Correlations**: According to the documentation, what are correlations between two columns of data (of your chosen set) and how should one interpret the output score? What are *high correlations* and what are *low-correlations*? You could also investigate using `pairs.panels()` to run some of your tests and you are to how to interpret these plots, if you employ them.

2. **Choose Six Columns From Your Data Set**: Perform correlation analysis over six columns of data using plots and textual outputs. Explain how to understand these results. Can you pick other columns which have stronger correlations to answer this particular problem?

3. **Justify Your Correlation Results**: Once you have found correlations in your data set, try to justify why the correlation exists. Remember that the data will not answer the *why* part of your work and so you may need to investigate academic articles to provide grounded theories to explain your results. You and your group are to use logical reasoning to give plausible reasoning to provide explanations about these correlation(s). Note: If no correlations were found in your data, then you could go back to pick new columns to run the test again or, you could explain why you think that correlations are not present.

4. **Five Answerable Questions**: Determine five basic questions to which you are able to provide inquest by using programming from *psych* package. For your questions, you are to try using the `t.test` or `lm` (linear model regression) to provide insight into your analysis. You are invited to use your slides, online resources or RStudio's online help to get you started (i.e., type `?lm` and `?t.test` in R for help).

5. ***p*-Values**: Your summary of each test from above will return a *p*-value. Explain how to understand the test from this statistical result.

**Important Details**

**Note: Please remember to include your name, as well as the names of your group members, on everything you submit for the class.**

**Required Deliverables**

1. File `analysis.r`: The R source code that you used to answer your questions.

2. File `report.md`: A Markdown report document listing all the outcomes of the above tasks. Namely, your report is to include the following.

   - The source of data,
   - The data descriptions,
   - The correlation analysis,
   - The logical reasoning behind the correlations,
   - The exploratory questions,
   - The statistical tests and their interpretations

Write your report so that it is clear which parts of your code and discussion addresses what particular question. Also, in your code, offer comments to state which code is for what exploratory problem.