

Name:

MATH 321: Homework 3

Due _____ : Turn in a hard copy, neat and stapled.

1. **EDA:** We are going to explore the 'Factors that Affect High School Completion Rates for 2013-2014' dataset. The goal is to summarize, describe and display information related to the included variables. In doing so, see if you can find any trends, patterns, interesting observations and try to relate it to the context. Try to find a narrative or a question to answer and use your EDA to investigate this.

This is intended to be a very open-ended assignment. You may explore as much or as little (as long as you meet the minimum requirements below) as you want. In addition you may do the work in Excel, R or both.

Files:

- If working in Excel, start with the 'high-school-data.xlsx' file on Canvas.
- If working in R, read the 'high-school-data.csv' file on Canvas into a .qmd file.

Requirements:

- Must create a minimum of two:
 - Set of descriptive statistics \rightarrow mean, st dev (sample), five-number summary, range and IQR.
 - Frequency and relative frequency table.
 - Histograms (of some kind: frequency, relative frequency, density).
 - Boxplots, if outliers are present extract / make note of them.
 - Scatterplots and correlation calculation (between a pair of variables).
- In addition:
 - Your exploration must include an analysis by region in some capacity (e.g. comparative boxplots, summary statistics by region, or anything else you can come up with).
 - These count towards the above requirements.

Submission: There will be two parts to the submission.

- R / Excel work (20 points): Submit the ORGANIZED Excel / .qmd file with all of the summaries and displays you created.
- (Informal) Write-Up (10 points): Based on the narrative you explored, copy the relevant summaries and displays from your Excel / R work into a word document.

Then write a summary of your findings, relating it back to the context of the data (this may include unanswered questions that perhaps more / different data would be needed to explore, etc.)

(A bulleted-list is perfectly fine, we will say 4 bullets minimum :)

Select answers

1. Answers may vary