

Name: _____

This is a closed-notes, email, lab assignments, lecture handouts, R Scripts exam.
The only outside documents you can use on the exam are the textbook *R Cookbook*, Google searches, and 3-page long cheat sheet that you prepared.

Notes: ~~~~~

Use R for all questions

- Create a new folder on your desktop and name it **Exam2_your name**. Set this as your working directory in RStudio.
 - **Open a blank Word document** and save it as **Exam2_Word_your name.docx**.
 - **Open RStudio**, open a blank source file (R Script) to work in, and make sure all History entries are cleared before you start your work on the exam questions.
 - In the R Script, add Exam-2 (by your name) as a comment line, and clearly label your answer to each question and part.
 - When recreating a plot, you will be graded on how accurate your recreation is. If you cannot recreate an aspect, submit your best approximation to be graded.
- ~~~~~

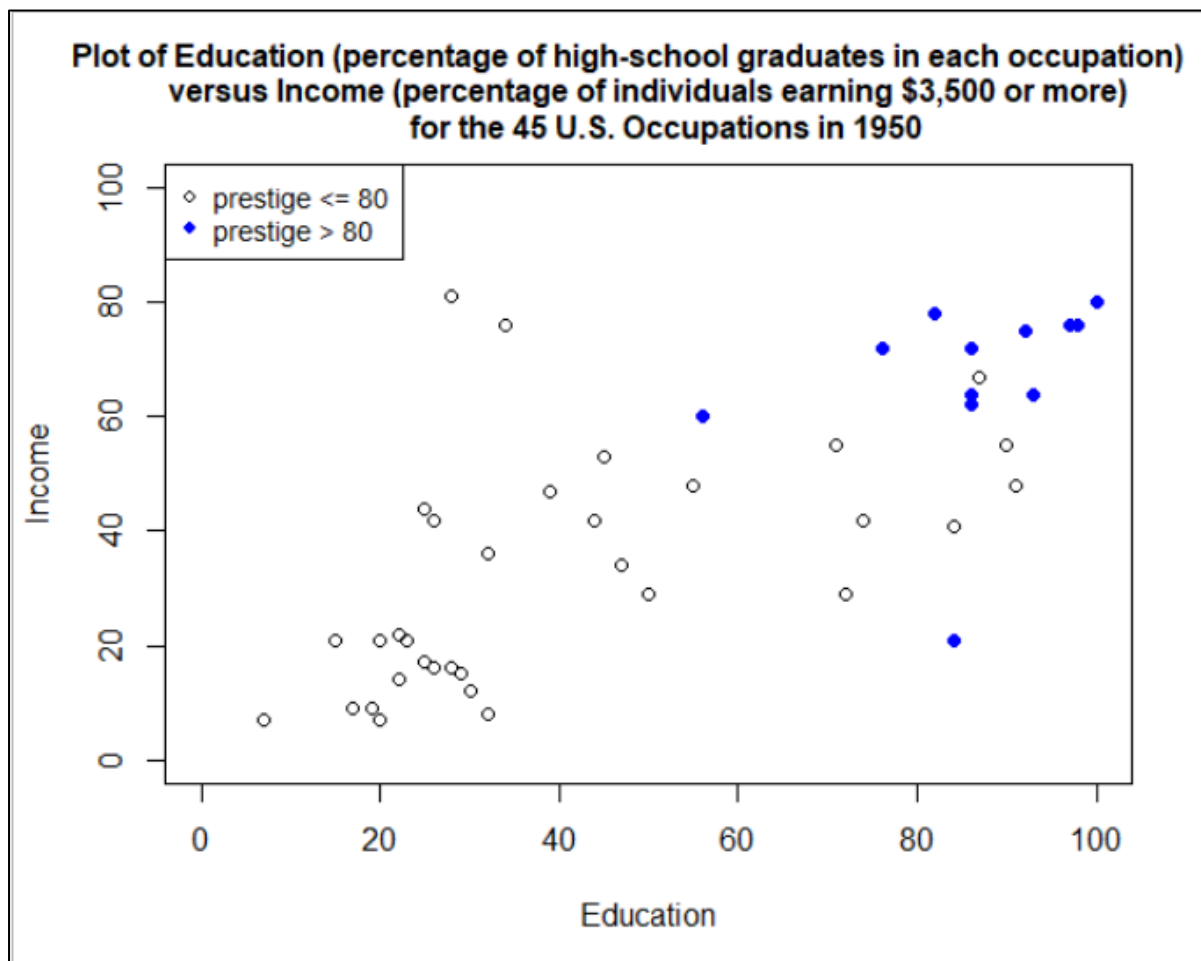
1. (17 pts) Go to the Moodle Course page. From **Data, Script, and other R Files** folder under **Course Documents**, download the Excel file **StuSurvey.xlsx** and save it into your **Exam2_your name** directory. (Hint: Don't forget to run **library(openxlsx)** command first)
 - a) Read this Excel file into an R data frame, naming it **Survey**.
 - b) Write and run the R command that will list the first eight rows and the first 10 variables of **Survey**.
 - c) Write and run the R command that will list the **CreditHr** and **Work** values for the 15th and 50th rows of **Survey**.
 - d) Write and run the R command that will subset the male data, naming the new data frame **MSurvey**.
 - e) The **Age** variable is recorded in months. Write and run the R command that will subset the dataset for all the students (not just males) who are younger than the median age, naming it **newSurvey**.

2. (20 pts.) Install and load the user-contributed package **car**. In that package there is a data frame called **Duncan**, which provides historical data on perceived job prestige in 1950. Type **?Duncan** and take a good look at the documentation on this data set. Then, do the following:

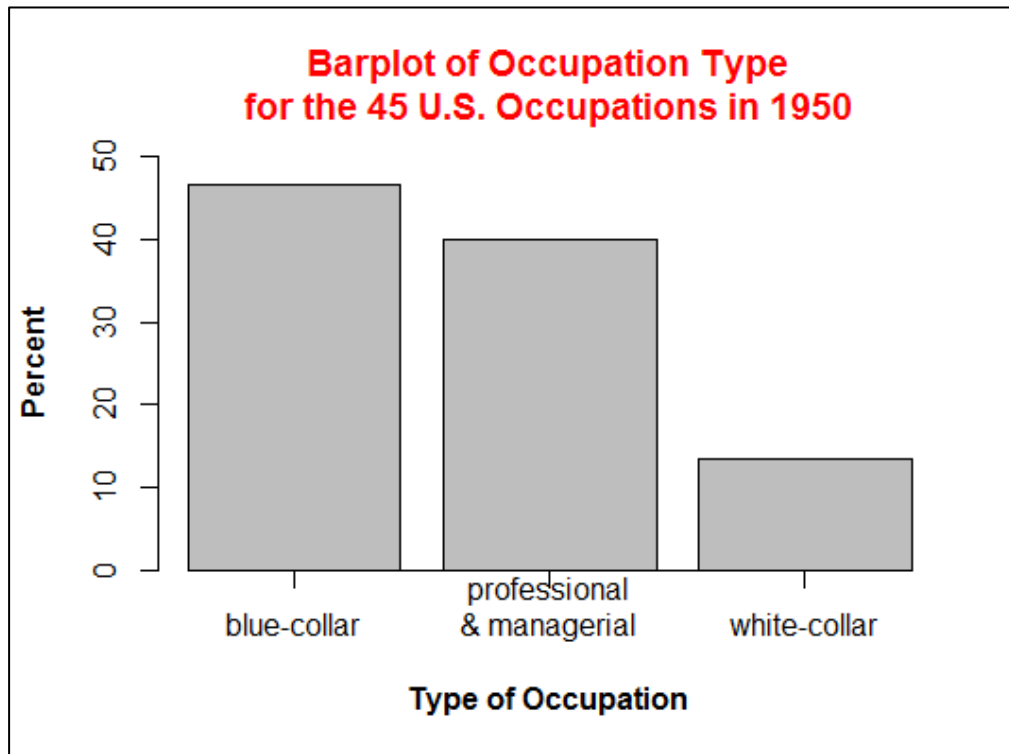
- a) Write and run R code that will plot **education** on the x-axis, and **income** on the y-axis, with both x- and y-axis limits fixed to be [0, 100]. Provide a meaningful title and appropriate axis labels. For jobs with a **prestige** value of less than or equal to 80, use a black \circ as the point character, and for jobs with **prestige** greater than 80, use a blue \bullet . Include a legend.

Copy your plot to the Clipboard and then paste it into your Word file.

Hint: You are expected to produce the plot given below



- b) Using the **Duncan** data frame, determine the command(s) that will recreate the following plot. Then, copy your plot to the Clipboard and paste it into your Word file.

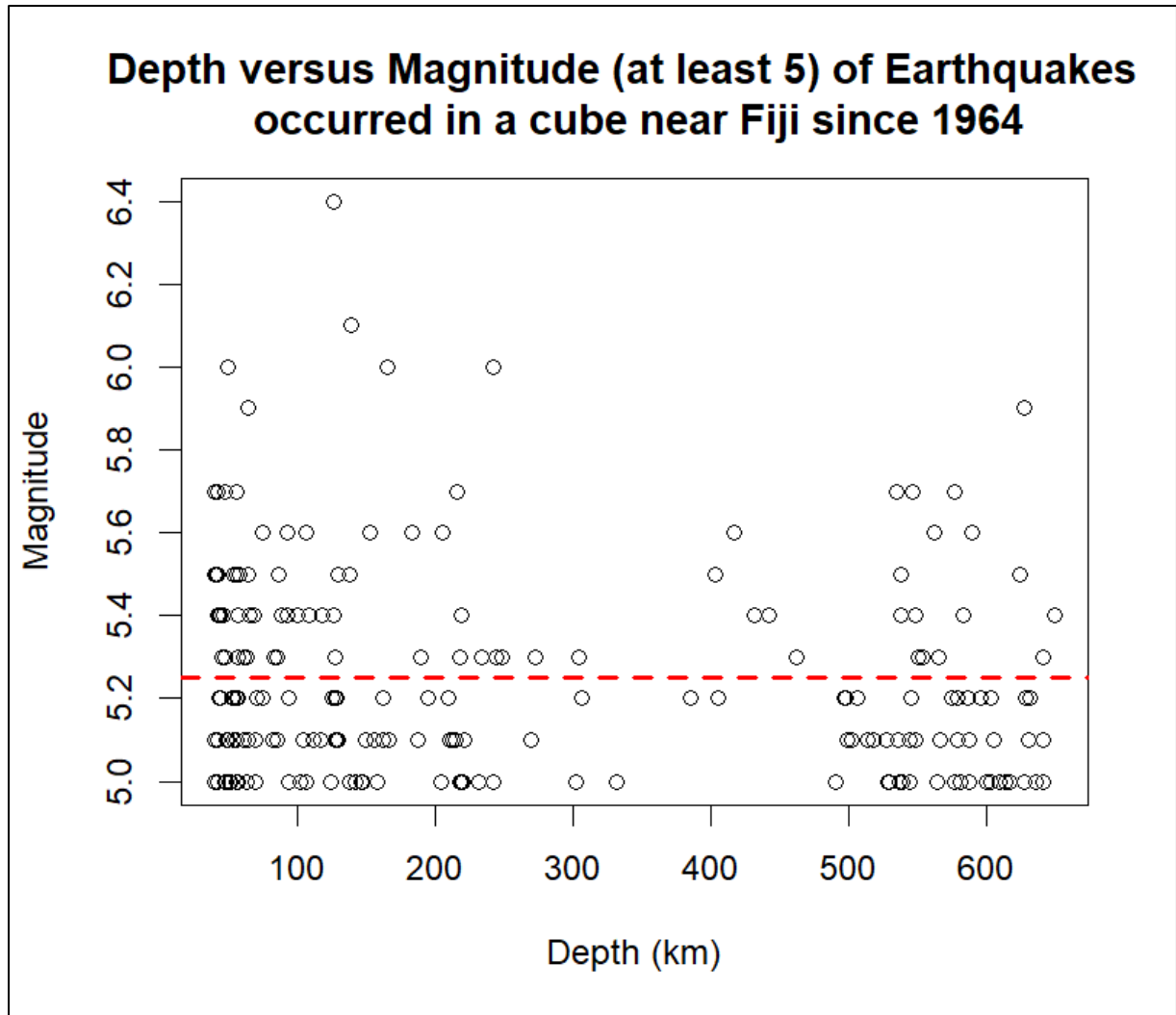


3. (25 pts.) The data frame **quakes** is in R's built-in **datasets** library. Access this object and view the corresponding help file to learn about this dataset. Then, do the following:
- Using the **quakes** data, write and run a single line of R code that will result in summary statistics (the five-number summary along with the mean) of the magnitude (**mag**) of the seismic events off the coast of Fiji that occurred at a depth of less than 400 km.
 - From all **quakes** data, select only those records that correspond to a magnitude (**mag**) of greater than or equal to 5 and write them to a table-format file called **q5.txt** in your **Exam2_your name** directory. Use delimiting character of ; and do not include any row names.
 - Read the file back into your R workspace, naming the object **q5.dframe**.
 - Using the data frame **q5.dframe**, the number of **stations** for the highest magnitude earthquake and report the result here: _____

e) Using the data frame `q5.dframe`, produce a plot of **depth** (on x-axis) versus **magnitude** (on y-axis).

- Be sure to add a meaningful title to your plot and label the axes.
- Add a dashed red line to your plot marking the mean magnitude.
- Copy your plot to the Clipboard and then paste it into your Word file.

Hint: You are expected to produce the plot given below



4. (20 pts.) Consider the data frame **chickwts** from R's built-in **datasets** library.

- a) Using **with()** and **tapply()** functions in a single line of R code, find the mean weights of the chicks for each feed type.

The mean weight of chicks is the highest with the _____ feed type and it is _____ grams.

The mean weight of chicks is the lowest with the _____ feed type and it is _____ grams.

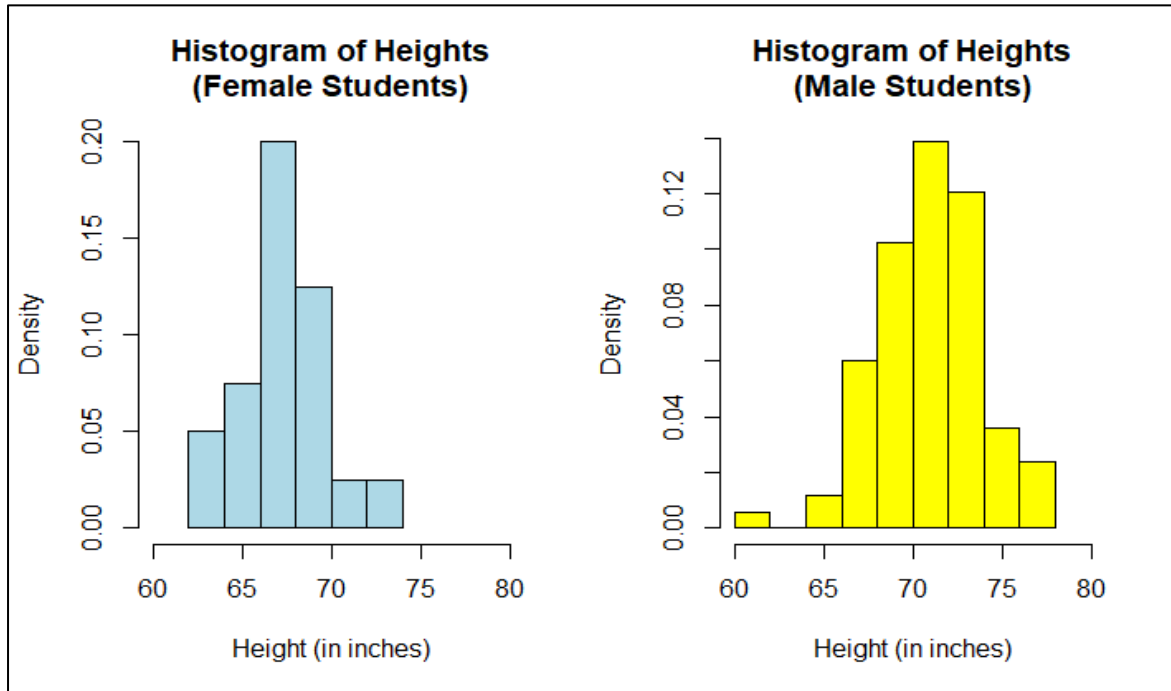
- b) Write and run the R commands that return the following table for the **feed** (the feed type) variable.

	Count	Percent
casein	12	16.90
horsebean	10	14.08
linseed	12	16.90
meatmeal	11	15.49
soybean	14	19.72
sunflower	12	16.90

5. (18 pts.)

- a) Write and run the R command that saves all of the objects that you have created in this session into a file **all.R**
- b) Write and run the R command that lists all the objects in your workspace.
- c) Write and run the R command that removes all the objects from your workspace. Then, check to make sure there are no objects remaining in your workspace.
- d) Write and run the R command that retrieves all the objects from **all.R** file. Then, check to make sure all objects are back in your workspace.

- e) Using the **Survey** data frame, determine the commands that will recreate the following graph for the **Ht** variable. Then, copy your plot to the Clipboard and paste it into your Word file. (*Hint: Both histograms have densities on y-axis, and they are displayed side-by-side in the same graph region*).



Save your **Exam2_Word_your name.docx** and **Exam2_fillable.pdf** files.

Save your RScript naming it **RScript_your name**.

Email the professor these five files at sgazioglu@mtech.edu :

Exam2_your name.pdf

RScript_your name

all.R

q5.txt

Exam2_Word_your name.docx

Have 'Stat435 – Exam2' in the subject line of the e-mail.