CS215 Chapter5E HW – Graphical Visualizations with Die Simulations Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Data was collected from two secondary schools in Portugal (denoted ‘GP’ and ‘MS’) in an attempt to determine what variables affects student grades. A description of what data has been recorded can be found below.

NOTE: The Portuguese system assigns each student a score from 0-20, but I have put these on a % scale (0, 5, 10, …100) so as to make this consistent with our American scale. You will see students’ final grades in the G3 column/variable below.

''' Attributes for both student-mat-posted.csv (Math course) datasets:

0 school - student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)

1 sex - student's sex (binary: "F" - female or "M" - male)

2 age - student's age (numeric: from 15 to 22)

3 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)

4 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)

5 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)

6 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)

7 activities - extra-curricular activities (binary: yes or no)

8 nursery - attended nursery school (binary: yes or no)

9 higher - wants to take higher education (binary: yes or no)

10 romantic - with a romantic relationship (binary: yes or no)

11 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

12 freetime - free time after school (numeric: from 1 - very low to 5 - very high)

13 goout - going out with friends (numeric: from 1 - very low to 5 - very high)

14 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

15 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

16 health - current health status (numeric: from 1 - very bad to 5 - very good)

17 absences - number of school absences (numeric: from 0 to 93)

18 G1 - first period grade (numeric: on their scale, it's from 0 to 20. I have converted it to 0 to 100.)

19 G2 - second period grade (numeric: on their scale, it's from 0 to 20. I have converted it to 0 to 100.)

20 G3 - final grade (numeric: on their scale, it's from 0 to 20. I have converted it to 0 to 100.)'''

Attached with this homework is a .py file which contains the following.

1. The actual data recorded is in the school list in the .py attached with this homework.
2. A descriptive\_stats function: Shows the mean, median, and standard deviation of a data set and

also gives the percentage of points falling one and two

std deviations about the mean.

1. A create\_freq\_barchart function, similar to the one you used in your lab

The Directions for your HW are on the back.

You will analyze this data with barcharts and descriptive stats so as to learn if the distribution of student grades appear to be affected by some of the above variables. You will use code to analyze the data and write your results in a word document.

1. You must answer these questions:
   1. Do the distributions of grades appear to be similar/different for those whose studytime values are different? If different, how so?
   2. Do the distributions of grades appear to be similar/different for those who are/are not in a romantic relationship?
   3. Many studies show that whether or not a student was in pre-K in the USA deeply affects student performance. So we may consider: Do the distributions of grades appear to be similar/different for those who did/did not attend nursery school?
2. Ask two more questions of your own. Find a relationship in which there appears to be a difference in grade distributions.