VARIABLES

ID (incremental)

SEX (M, F)

AGE (in years)

WEIGHT (in kgs)

HEIGHT (inches)

DIAGNOSIS/CAUSE (diabetes, hypertension, glomerulonephritis, undetermined, other)

COMORBIDITIES (yes, no)

TREATMENT PERIOD (in years)

DIALYSIS CHARACTERISTIC PERCENTAGE (weekly, high flux, short term CVC, long term CVC)

PATIENT’S FINAL SITUATION (haemodialysis, transferred, transplant, died)

DATA CLEANING, ANALYSIS AND VISUALIZATION GUIDE.

Session – R Programming

1. Examine the EXCEL dataset.

2. Import the EXCEL dataset to R data set dialysis.

3. Create a permanent R data set called dialysis and store it in your directory.

4. How many observations are in the data set? How many variables? Is the variable weight numeric or character? Is the variable sex numeric or character? What command did you use to answer these questions?

5. Clean the comorbidities, and diagnostic\_cause variables.

6. How many missing values does the weight variable have? Clean the weight variable. (missing values and values less than 50 change to 70)

7. Create the new variable BMI. The formula is:

BMI = weight (kg) / [height (m) \* height (m)]

Conversion: meters = 0.0254 \* inches

8. What is the average BMI value? What is the median BMI value?

9. Create a new variable to identify the BMI categories and call it BMIcat:

- underweight (20.4 or under) ("under")

- normal weight (20.5 - 27.9) ("normal")

- overweight (28 or greater) ("over")

10. How many persons are still in haemodialysis? How many men are dead? How many women over the age of 45 have comorbid conditions? What is the percentage of men over the age of 60 have glomerulonephritis?

11. Clean the sex variable. Create a new variable called gender which has values “Male” and “Female”.

12. Produce a pie chart for the variable gender as a visual representation.

13. Produce a bar chart for the variable diagnostic\_cause as a visual representation.

14. Produce a scatter plot for the variables age by weight. Does there appear to be an association between the two variables?

15. Produce a boxplot visual presentation of weight by comorbidities.

16. Produce a bar chart for the categorical variable BMIcat that was created in question 9.