Exercise 1 – Transforming and analyzing mental health data

The aim of this exercise is to use the basic R functionality covered in the training content to transform and analyze a mental health data set. In the exercise you will import data, subset the data, conduct tabulation and create a for loop to analyze patient length of stay.

1. Set the working directory

--> Check

2. Import the data set PD_Data.csv
    - Dataset is posted on PD_Data_R_Training.csv

--> Check

3. Calculate length of stay for all patients (you will need to convert the dates using as.date)

--> Check

4. Subset the data by the four care settings (setting column)

--> Check

5. Table ward/team and ICD-10 code for each subset

--> Check

6. Table ICD-10 code by cluster for each subset

--> Check

1. Create a for loop to calculate the total and mean length of stay for each unique patient (here is a quick tutorial on how to write a for loop: https://rforhealthcare.org/for-and-while-loops/)

    1. Create data frame with just client ID and length of stay

--> Check

    2. Remove NA values

--> Check

    3. Create a vector of unique client ID’s

--> Check

    4. Create vectors of zero for the total and mean values

--> Check

    5. Begin the for loop

--> Check

    6. Subset for each client ID

--> Check

7. Calculate the total length of stay for each client ID

--> Check

8. Calculate the mean length of stay for each client ID

--> Check

9. End the for loop

--> Check

10. Bind the unique client ID, total length of stay and mean length of stay vectors

--> Check

Example code for the completion of this exercise can be found here. Don't worry if you have achieved the same outcome a slightly different way, there are many ways to do the same task.

Exercise 2 – plotting and statistically describing mental health data

In this exercise we will build on the analysis you undertook in exercise 1 to create plots, descriptive statistics and conduct a statistical test on mental health service use data.

1. Plot length of stay as a histogram

--> Check

2. Describe length of stay

--> Check

3. Describe length of stay by cluster

--> Check

4. Subset the data by cluster and NA cluster values (is.na)

--> Check

5. Create histograms for each cluster subset

--> Check

6. Conduct t-tests comparing each cluster subset

--> Check

7. Create a box plot of all three cluster subsets

--> Check

Example code for the completion of this exercise can be found here. Don't worry if you have achieved the same outcome a slightly different way, there are many ways to do the same task.