# Lab 3: Week 3- Distributions

Things To Do

1) Load the blank Lab 3 template, once loaded make sure you save it with your name

2) Chunk 1: Load Libraries and Data

1.1) Load the packages “ggplot2”, “tidyverse" and “datasets”

1.2) Load the dataset “sleep”

1.3) Using a comment describe the structure `str()` of the ‘sleep’ dataset

3) Chunk 2: Histograms

2.1) Using ‘ggplot’, ‘geom\_histogram()’, ‘gg\_title’ plot a histogram of the sleep dataset’s variable “extra” using the default bins setting and title each plot based on the number of bins displayed.

2.2) Now modify the `bins` argument from 30 (default) and create new plots, (change

the title to reflect the number of bins being displayed)

bins=5

bins=10

bins=20

bins=50

2.3) Using a comment describe what happens when you change the bin size?

4) Chunk 3: Density Plots

3.1) Create a density plot for the variable “extra” in the “sleep” dataset

3.2) modify the x axis to range from -10 to 10

3.3) Create a new density plot but color the density plot by “group” with a 50%

transparency

5) Chunk 4: Random (Standard Normal) Distributions

4.1 Create an object called “extra” of a normal distribution with a 20 observations a

mean of 0 and a standard deviation of 1.

4.2. Convert norm from a list to a data.frame (?as.data.frame)

4.3 Plot it as a density plot

4.4 Change the number of observations to 10 and plot it as a density plot. Using a

comment, how does the distribution change?

4.5. Change the number of observations to 100 and plot it as a density plot. Using a

comment, how does the distribution change?

4.6 Use `shapiro.test` to test for normality in this distribution? What do the results suggest? Are the

data normally distributed?

6) Chunk 5: Data Wrangling and Joining

5.1) Create a new object called “norm.table”; Add two columns to the `extra` dataframe

and name them “group” and “ID” fill the cells with the value “RANDOM”

5.2) Bind the ‘norm.table’ object with the ‘sleep’ data in a new object called ‘new.df’ (?rbind)

5.3) Plot the ‘new.df’ object as a density plot, colored by “group” with a 50% transparency

5.4) Use the `filter` function to select only the observations for group 1; are these data normally

distributed?

5.5) Use the `filter` function to select only the observations for group 2; are these data normally

distributed?

7) Chunk 6: Checking for `kurtosis` and `skewness`

6.1. load the library `(e1071)` which has the functions kurtosis and skewness? What are the arguments

in these functions?

6.2 Compare the Kurtosis and skewness between the random normal distribution you generated and

group 1 and group 2.

6.3 For kurtosis and skewness set the type=2 for unbiased normal data.

8) Chunk 7: Fancy plotting

Do any five of these things to the version of the plot completed in chunk 5:

 change the default color of the plot to different color using ‘scale\_fill\_manual’

 change the x label to “Effect on hours of sleep”

 change the background theme to black and white ‘theme\_bw()’

 add a vertical line at the mean value ‘geom\_vline()’

 change the label on the color key ‘labs’

 change the position of the key from the right to the top or bottom

 add a title to the plot ‘ggtitle’

 annotate the plot with some text (e.g. describing the vertical line) ‘annotate’

 change the figure size to 4 in X 6 in (markdown chunk modification)

 change the output resolution of the figure to 400 dpi (markdown chunk modification)