Arjun Krishnan

Day 08 Pre-class Assignment

For this week's assignment, I would like you to complete the exercise you began in class and prepare for the class on Monday.

1. Effect of sample size on correlation coefficients

You were given R code to visualize the effect of sample size on observed correlations by doing the following simulation:

- For each sample size n = {5, 10, 50, 100}:
  - Repeat 10,000: Calculate correlation coefficients of samples of two independent uniformly distributed variables between 0 and 1.
  - Plot a histogram of the 10,000 correlation coefficients, which describes the distribution of correlations under the null hypothesis.

As part of your assignment, extend this exercise to do the following:

- For sample size,
  - Mark the coefficient in each histogram that corresponds to α = 0.05.
  - Make a separate scatter plot of the pairs of data points with the three largest and smallest correlation coefficients.

2. Creating a lineup for visual inference

Write code to create a lineup plot to visually distinguish the "real" dataset from 19 other randomized versions of this dataset.

- Create a pair of variables with 20 observations each that have a chance of being correlated with each other:

```
xunif <- runif(20)
yunif <- xunif + rnorm(20, sd = 0.6)
```

- Now, create 19 more versions of this same exact dataset but, each time shuffling the `yunif` variable, keeping `xunif` the same. Hint: use the `sample` function.
- Generate a random 5-digit dataset identifier for each dataset. Hint: `paste(c(sample(1:9,1), sample( 0:9, 4, replace=TRUE )), collapse="")`.
- For ease, aggregate all these twenty datasets – one real & 19 shuffled – into a single dataframe with 3 columns – `<dataset_id> <x_value> <y_value>`.
- Plot all twenty plots in a randomized order – one scatterplot per dataset – and comment about whether you are able to spot the real scatterplot from the randomized ones.
- Check your visual inference by doing a statistical test:

```
cor.test(xunif, yunif)  # check the statistical significance of the correlation
```

3. Prepare for visualization challenges

No need to submit anything for this third part but I’m asking you to brush-up making the following plots: scatterplot, line graph, pie chart, boxplot, violinplot, barplots and error bars, changing axes to log scale, changing x- and y-axis limits, histogram and changing bin size, using color scale/palettes, adding secondary y-axis, adding colors to points/bars, changing shapes of points, etc.