is607 Week 5 Quiz
Prashant B. Bhuyan
==========

1. Data ***tidying*** is what Hadley Wicham calls structuring datasets to facilitate analysis.

2. The three properties of tidy data are:

A. Each type of observational unit forms a table.

B. Each variable forms a column.

C. Each observation forms a row.

3. The five most common problems with messy datasets according to Hadley Wickham are:

A. Column headers are values, not variable names.

B. Multiple variables are stored in one column.

C. Variables are stored in both rows and columns.

D. Multiple types of observational units are stored in the same table.

E. A single observational unit is stored in multiple tables.

4. The four verbs of data manipulation can each be described as:

A. Filter: To conditionally remove observations or subset.

B. Transform: To modify or add single or multiple variables.

C. Aggregate: To collapse many values into a single value.

D. Sort: To change the order of observations.

5. The function in base R used to filter in data manipulation is ***subset()***

6. Compared to plyr, dplyr is faster and is better for joining; however, it only provides tools for working with data frames.

7. plyr is very good for grouping and applying the split-apply-combine method to large data sets. plyr can advantage multiple processors and parallelize large data sets making it more efficient than the base R apply functions. plyr also provides output consistency as well as consistency in names and arguments. The base R apply functions are good for summarizing data but plyr is as well. Base R apply functions don't provide a progress bar for long operations like plyr which is a big minus. Finally, base R apply functions don't provide labels that are maintained across all transformations.

8. The tidyr function that corresponds to pivot in spreadsheets is ***gather()***. The tidyr function that corresponds to unpivot in spreadsheets is ***spread()***.