## Week 5 Quiz

### Reading Quiz “Tidy data”

1. Data \_\_tidying\_\_ is what Hadley Wickham calls structuring datasets to facilitate analysis.
2. The other two are:
   * + Each variable forms a column.
     + Each observation forms a row.
3. The five most common problems with messy datasets.
   * + Column headers are values, not variable names.
     + Multiple variables are stored in one column.
     + Variables are stored in both rows and columns.
     + Multiple types of observational units are stored in the same table.
     + A single observational unit is stored in multiple tables.
4. **Filter**: subsetting or removing observations based on some condition.

**Transform**: adding or modifying variables. These modifications can involve either a

single variable (e.g., log-transformation), or multiple variables (e.g., computing density from weight and volume).

**Aggregate:** collapsing multiple values into a single value (e.g., by summing or taking

means).

**Sort**: ordering the observations.

1. subset()

### Reading Quiz “Introduction to dplyr”

1. Advantages of dplyr over plyr
   * 1. dplyr is faster than plyr.
     2. plyr only provides tools for working with data frames. While as dplyr provides tools for other datatypes like arrays, vectors, and lists.

Disadvantages of using plyr over R base package

plyr provides a better thought out set of joins.

1. Advantages of plyr over basic R packages
   * 1. plyr provides a replacement for loops for a large set of practical problems.

The tools of plyr aim to eliminate this extra code and illuminate the key components of the computation and abstracts away from the details of the underlying data structure.

Disadvantages of plyr over basic R packages

plyr makes the strong assumption that each piece of data will be processed only once and independently of all other pieces This means that we cannot use these tools when

each iteration requires overlapping data (like a running mean), or it depends on the previous

iteration (like in a dynamic simulation).

### Reading Quiz: “tidyr”

1. gather() and spread()