Tidyverse training

1. Know your data
   1. glimpse()
   2. skim()
2. Count with your data
   1. distinct()
   2. count()
      1. adds a new column (n)
         1. same as group\_by() and summarize()
         2. doesn’t require ungroup() for calling mutate()
3. Cast column types
   1. parse\_number() and col\_number
   2. parsing dates
      1. parse\_datetime()
4. Select variables
   1. Select()
      1. starts\_with()
      2. ends\_with()
      3. contains()
5. Tame variable names
6. Rename variables to convention
   1. janitor package
      1. clean\_names()
7. Tame data is not the same as tidy data
   1. What is the difference?
   2. Tame but not tidy
      1. Rectangle, each columns contain only one variable, variable names are easy to work with, variable types match the values. But not tidy if observations are spread across columns and not on different rows. Wide data
   3. Tidy data tends to be long, more repetitive
8. gather()
   1. reshaping data from wide to long
9. separate()
   1. three arguments – data, col, into
10. unite()
    1. Opposite of separate
11. spread()
    1. opposite of gather()
    2. use when there is a column with a mix of variable names
    3. arguments – data, key and value
12. case\_when()
    1. Think of it as an ifelse that you can chain together and pipe