Midterm Exam Study Guide

1. Install and load a library. Access data from that library. Check the dimensions and structure of a dataset.
2. Check the data type of a column. Check the levels of a factor.
3. Get the maximum value of a column, and find data associated with that maximum value.
   1. Specific functions to know: max(), filter()
4. Use mutate to create a new column. Categorize a continuous variable into groups.
   1. Specific functions to know: mutate(), case\_when(), between()
5. Filter by multiple columns, and deal with missing values.
   1. Specific functions to know: filter(), distinct()
6. Create a table with data from multiple columns.
   1. Specific functions to know: either table() [base R] or count() [dplyr]
7. Filter rows and select columns to create a new dataframe.
8. Create a new, calculated variable.
   1. Specific functions to know: mutate()
9. Create a histogram and describe it.
   1. Specific function to know: ggplot()
10. Create a density plot with a log2 x axis scale.
    1. Specific ggplot arguments to know: scale\_x\_continuous(trans=”log2”)
11. Create a scatter plot with a log2 y axis scale.
12. Plot a basic linear model over a scatter plot.
    1. Specific functions to know: lm(), data\_grid(), add\_predictions(), ggplot()
    2. Note: data\_grid() and add\_predictions() are from the modelr library.