Stat 6620

Va7892

Surabhi Asati

Homework 1

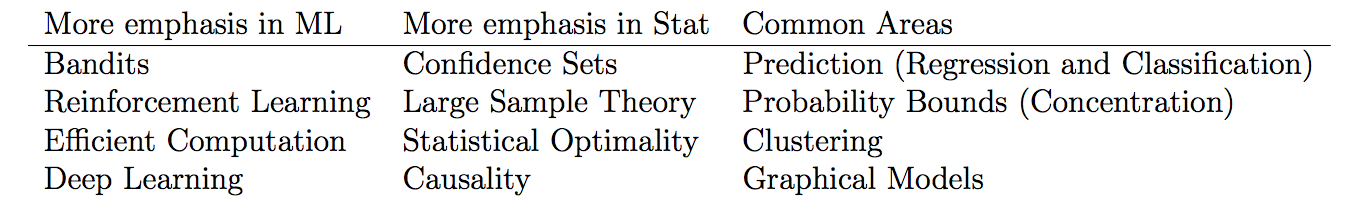
1. Do a google search on the following terms and develop a working definition of each.

Statistical Learning

Statistical learning theory is a framework for machine learning, drawing from the fields of statistics and functional analysis. Statistical learning theory deals with the problem of finding a predictive function based on data. The basic assumption made by statistical learning is that data with same properties (such as English articles, web data, image data, etc) have certain statistical regularity. Therefore, we can process them using probabilistic methods. For example, we can use random variables to describe features; we can use probabilistic distributions to describe the statistical regularity of data.

Statistical Machine Learning

Statistics and ML are overlapping fields. Both address the same question: how do we extract information from data? But there are differences in emphasis. In particular, some topics get greater emphasis than others. The term "statistical" in the title reflects the emphasis on statistical analysis and methodology, which is the predominant approach in modern machine learning. Here are some examples:



Machine Learning

Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases.

Predicitive Analytics

Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future.

Artifical Intelligence

Artificial intelligence (AI) makes it possible for machines to learn from experience, adjust to new inputs and perform human-like tasks. Most AI examples that you hear about today – from chess-playing computers to self-driving cars – rely heavily on deep learning and natural language processing. Using these technologies, computers can be trained to accomplish specific tasks by processing large amounts of data and recognizing patterns in the data.

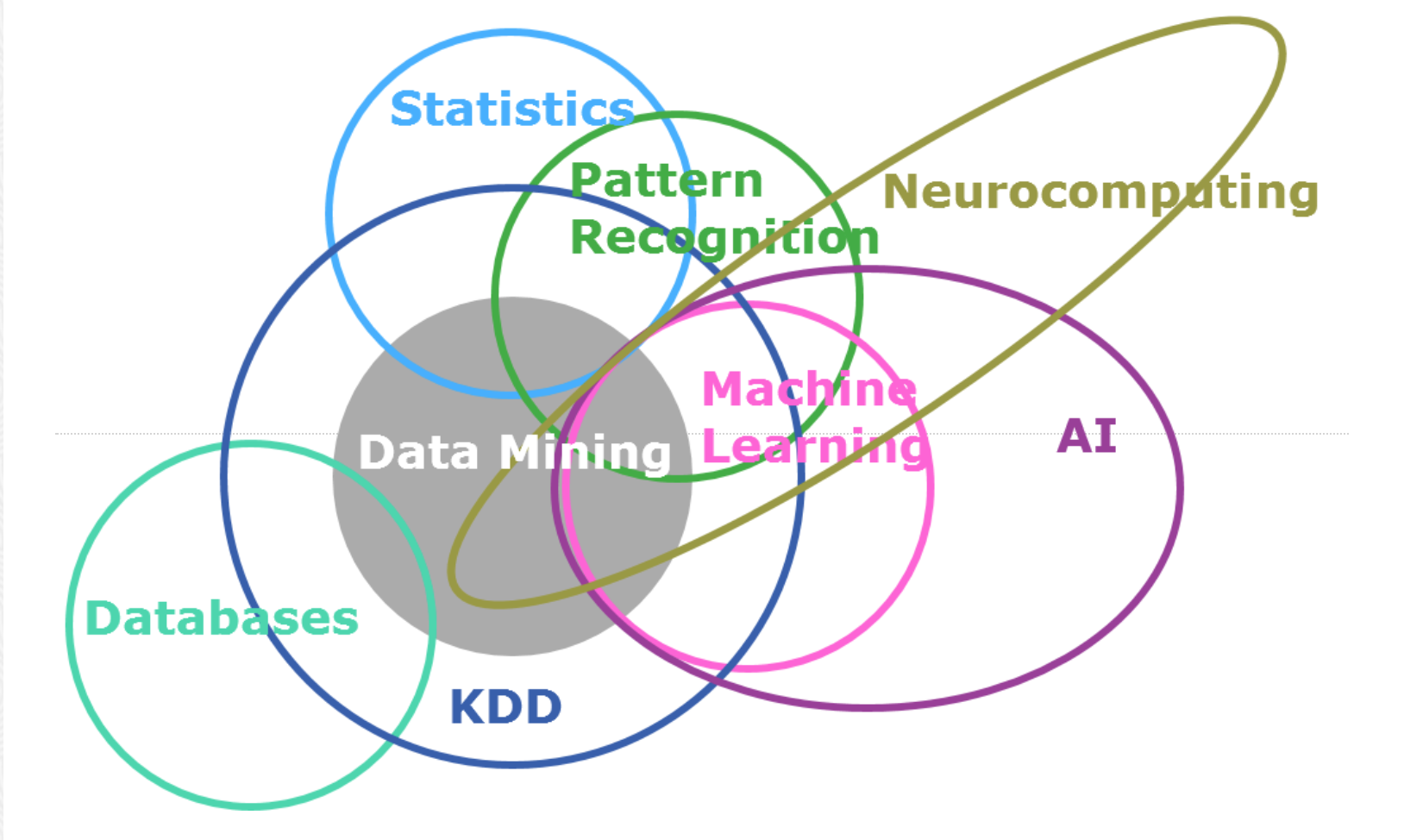
Deep Learning

Deep learning is a type of machine learning that trains a computer to perform human-like tasks, such as recognizing speech, identifying images or making predictions. Instead of organizing data to run through predefined equations, deep learning sets up basic parameters about the data and trains the computer to learn on its own by recognizing patterns using many layers of processing. Deep Learning is used by Google in its voice and image recognition algorithms, by Netflix and Amazon to decide what you want to watch or buy next.

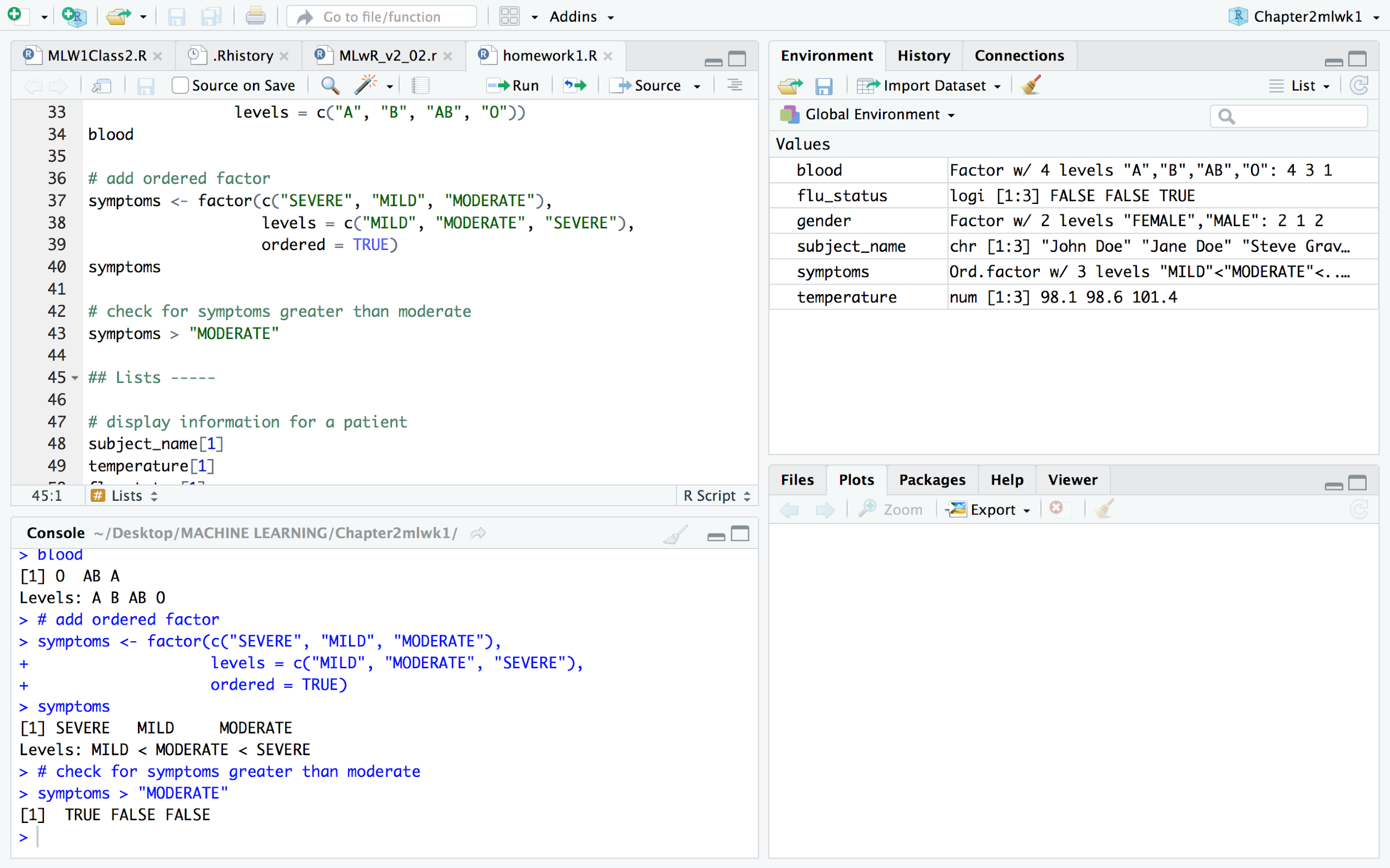
While ML is often described as a sub-discipline of AI, it’s better to think of it as the current state-of-the-art – it’s the field of AI which today is showing the most promise at providing tools that industry and society can use to drive change.

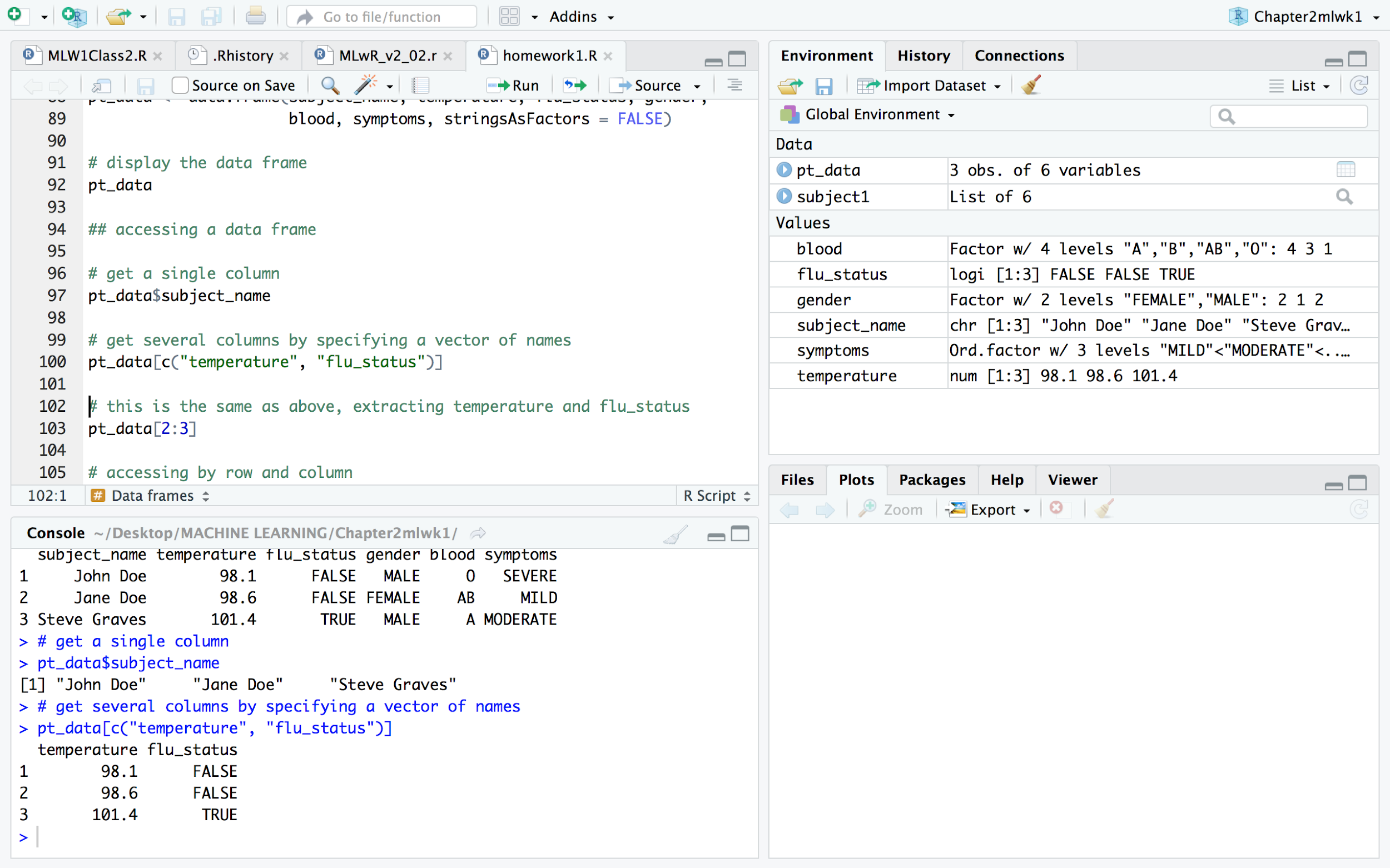
In turn, it’s probably most helpful to think of Deep Learning as the cutting-edge of the cutting-edge. ML takes some of the core ideas of AI and focuses them on solving real-world problems with neural networks designed to mimic our own decision-making. Deep Learning focuses even more narrowly on a subset of ML tools and techniques, and applies them to solving just about any problem which requires “thought” – human or artificial.

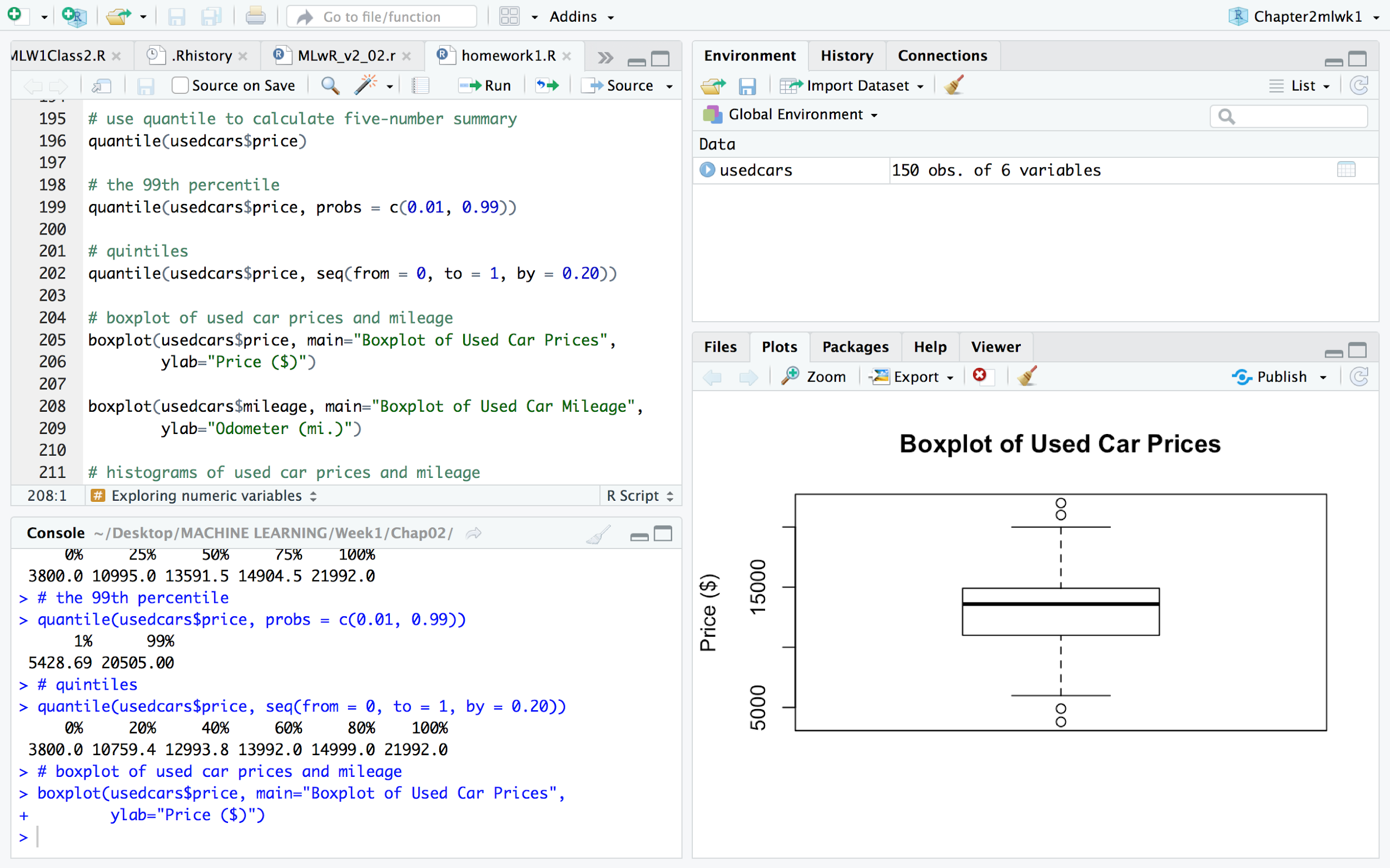
Here is an interesting Venn diagram on the coverage of machine learning and statistical modeling in the universe of data science:

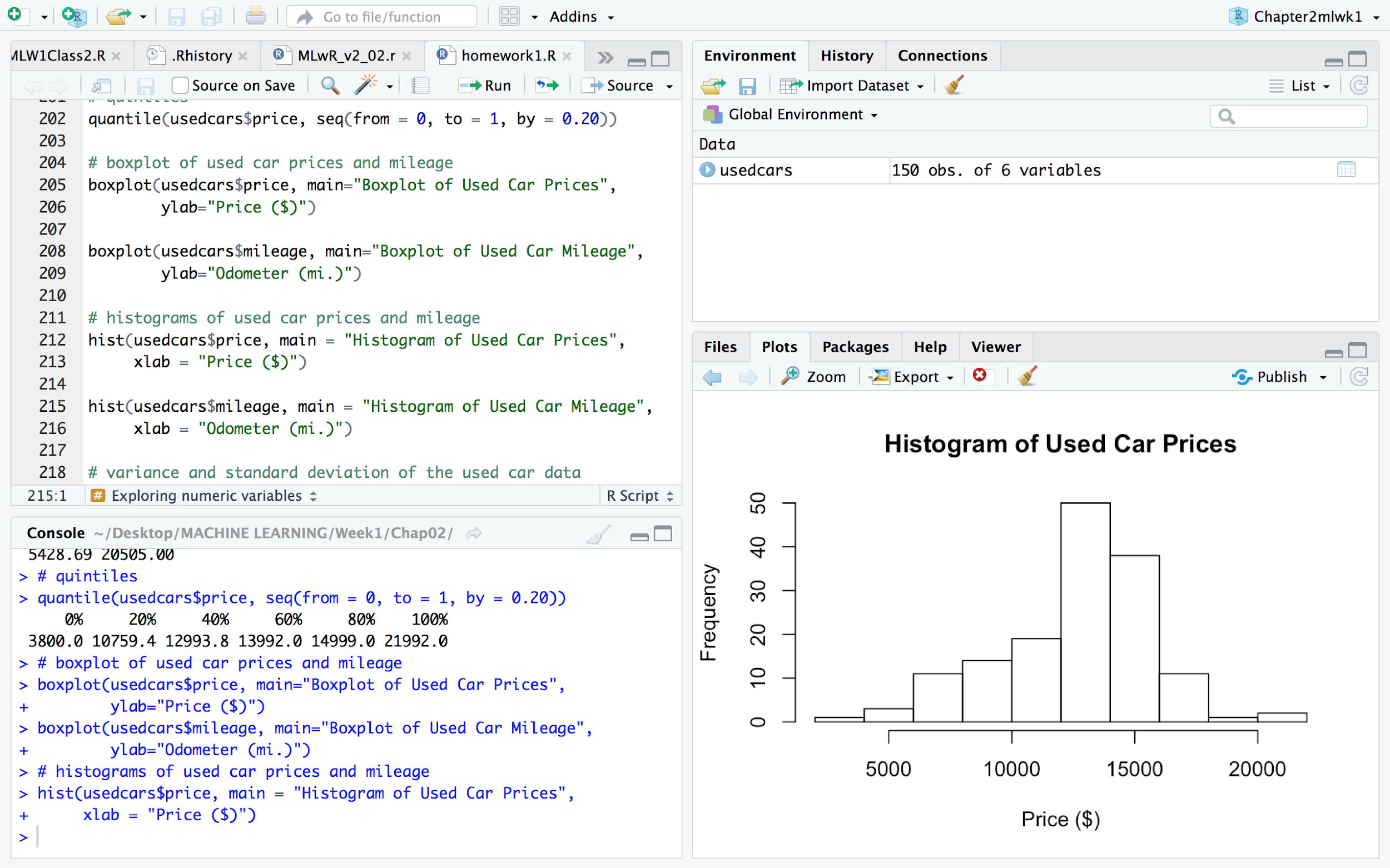


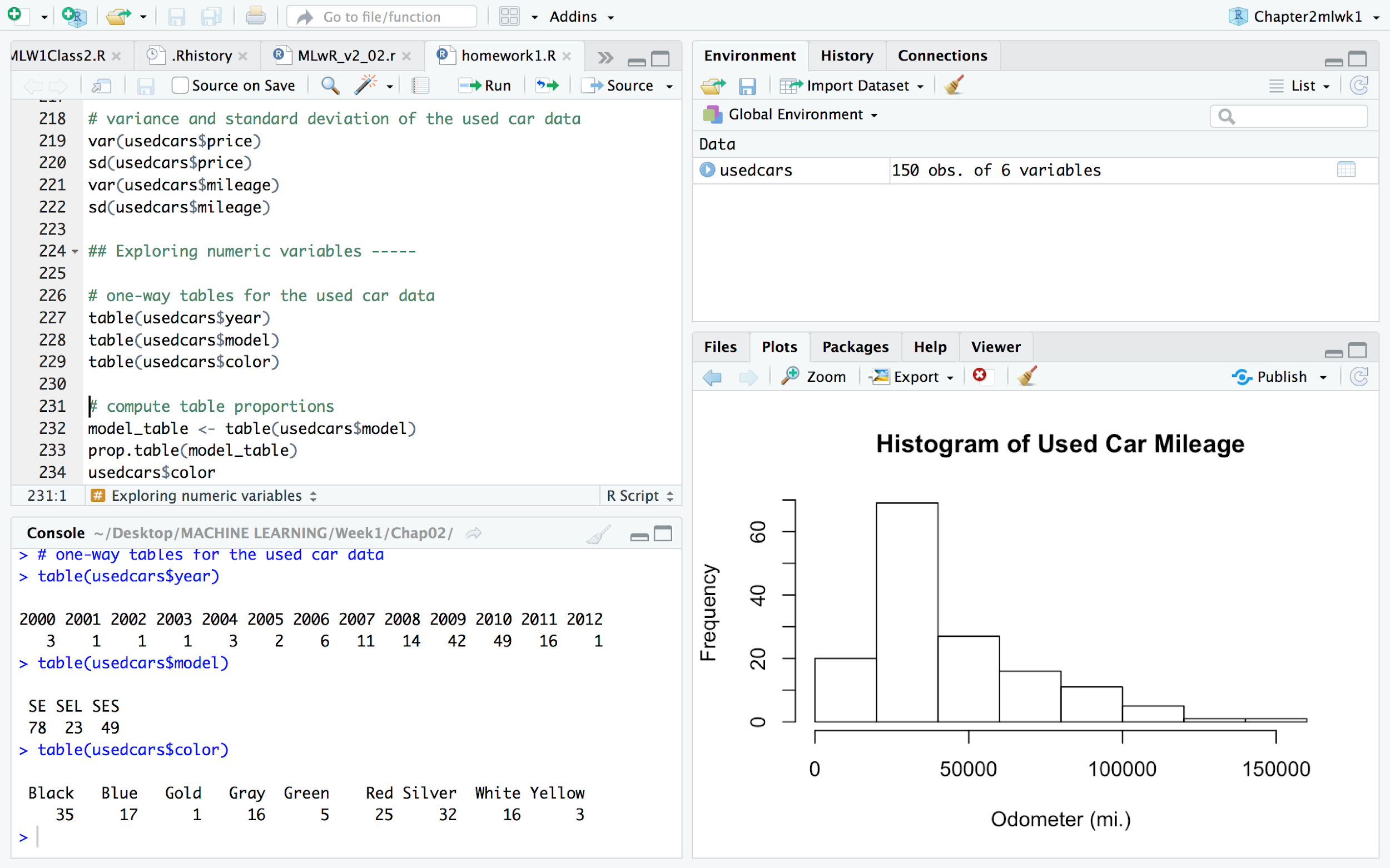
2. Run all of the code from Chapter 2 to become familiar with R. (If you have experience with R, this will get you familiar with the code from the author.) Show some of the relevant output from R and discuss what you have learned from the data.

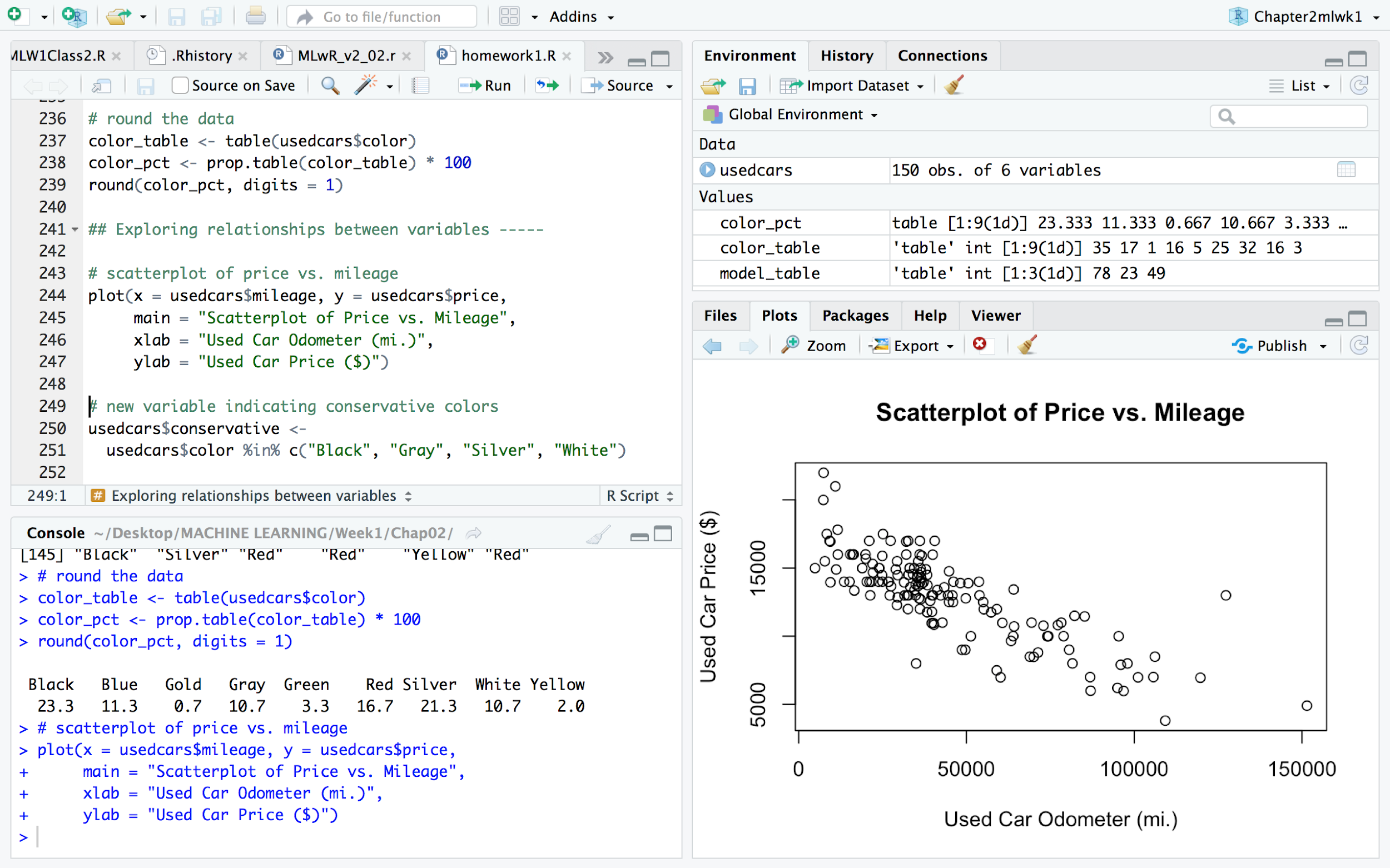


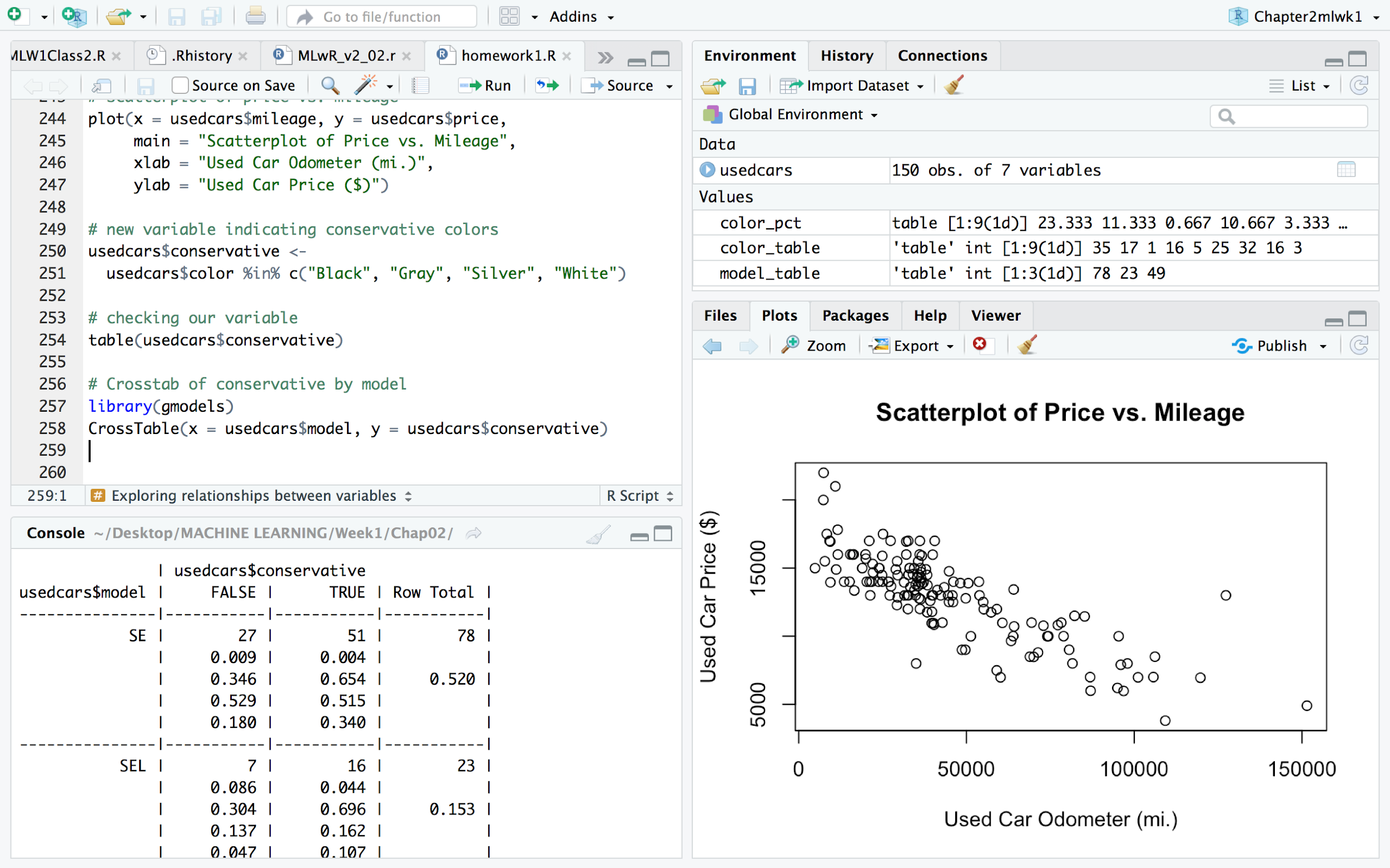


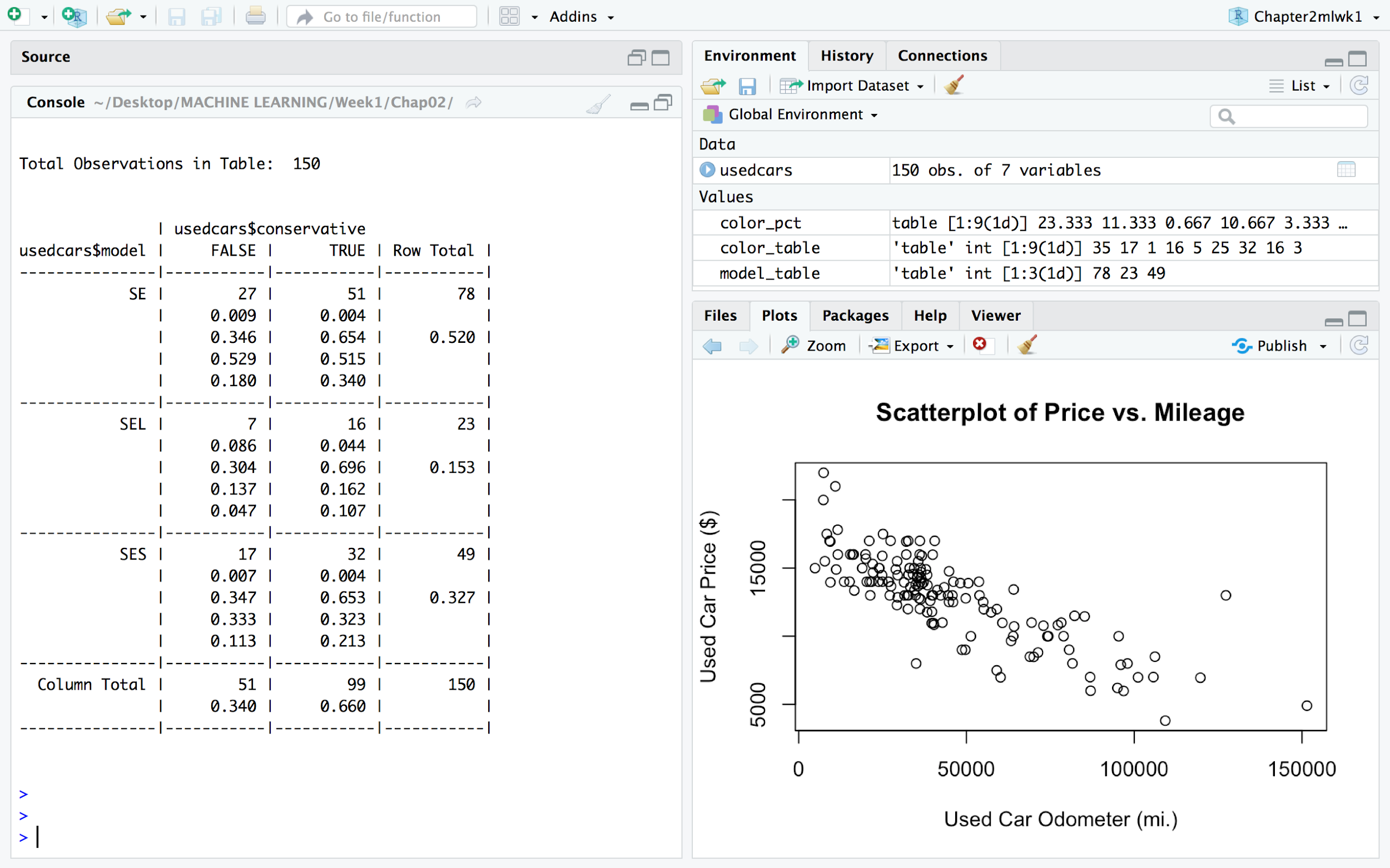












Learnings:

1. Basics of R using used car prices
2. Relationship visualization with scatterplots
3. Centre and spread of some numeric variables, crosstables etc.

3. Download this book and become familiar with the materials on the websites.

An Introduction to Statistical Learning

What do the authors of the Introduction book say about Statistical Learning?

(Optional: If you are interested start watching the 15 hours of video.)

Author view on statistical learning:

Statistical learning refers to a set of tools for modeling and understanding complex datasets. It is a recently developed area in statistics and blends with parallel developments in computer science and, in particular, machine learning. The field encompasses many methods such as the lasso and sparse regression, classification and regression trees, and boosting and support vector machines.

With the explosion of “Big Data” problems, statistical learning has become a very hot field in many scientific areas as well as marketing, finance, and other business disciplines. People with statistical learning skills are in high demand.