
MATH1401

Fall 2021

Lecture 7

Visualization – Part I

Class Checklist

- **Lab 2 – Due Date** : Thursday 9/9 – 9 PM
 - Graded Questions : 1.1, 2.1,2.1.1,3.1-3.6, 4.1-4.2
 - **HW - 2 Due Date** : Tuesday: 9/14 – 9 PM
 - Graded Questions:1.1-1.3, 2.1-2.5, 3.1-3.5, 4.1-4.2, 6.1-6.7
 - **Quiz 5** – Thursday: 9/9 – Covers Chapter 6
 - **Quiz 6** – Tuesday: 9/14 – Covers Chapter 7
-

Table Review

Lecture 6 Checklist

- Use **table.where(column,are.condition)**
 - Numerical Data
 - Categorical Data
-

Manipulating Tables

- `t.where(column, value)` keeps all rows for which a column's value equals some particular value
 - `t.where(column, are.condition)` keeps all rows for which a column's value satisfies a condition
 - `t.column('columnName')` returns column as an array
 - `t.num_rows` finds number of rows
-

Types of Attributes

All values in a column of a table should be both the same type **and** be comparable to each other in some way

- **Numerical** — Each value is from a numerical scale
 - Numerical measurements are ordered
 - Differences are meaningful
 - **Categorical** — Each value is from a fixed inventory
 - May or may not have an ordering
 - Categories are the same or different
-

Census Data

The Decennial Census

- Every ten years, the Census Bureau counts how many people there are in the U.S.
- In between censuses, the Bureau estimates how many people there are each year.
- Article 1, Section 2 of the Constitution:
 - “Representatives and direct Taxes shall be apportioned among the several States ... according to their respective Numbers ...”

(Demo)

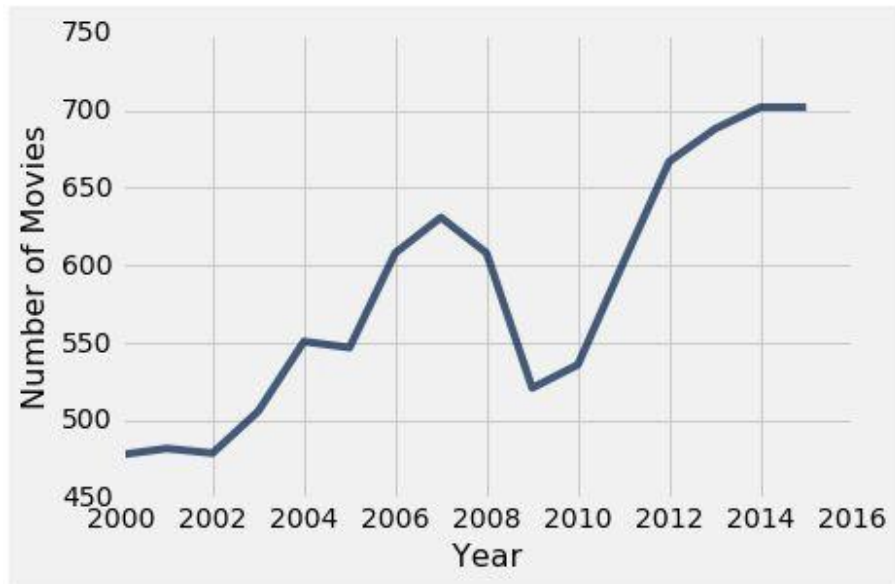
Lecture 7 Checklist – Chapter 7, 7.1

- Identify Numerical Data
 - Identify Categorical Data
 - Plot numerical data with Line plot, Scatter Plot
 - Plot categorical data with bar graph
-

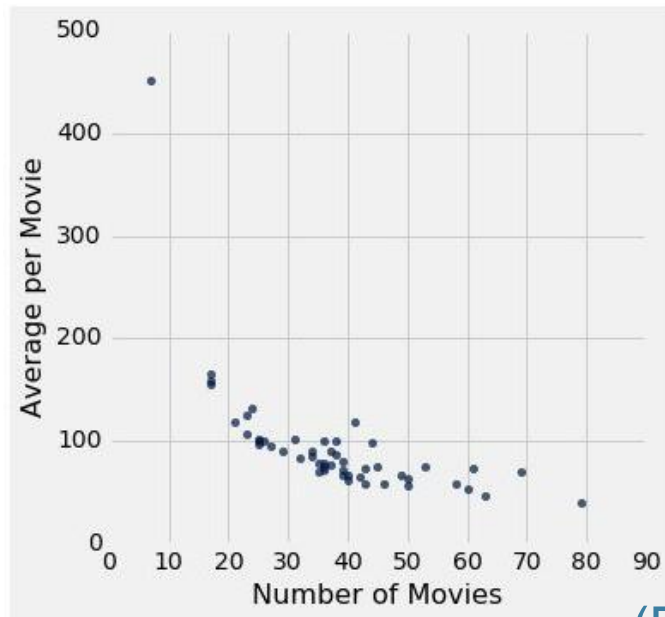
Numerical Data

Plotting Two Numerical Variables

Line graph: `plot`



Scatter plot : `scatter`



(Demo)

When to use a line vs scatter plot?

- Use line plots for sequential data: if...
 - ...your x-axis has an order
 - ...sequential differences in y values are meaningful
 - ...there's only one y-value for each x-value
 - Usually: x-axis is **time** or **distance**
 - Use scatter plots for non-sequential data
 - When you're looking for associations
-

Categorical Data

(Demo)

How Do You Generate This Chart?

