

# BC COMS 1016: Intro to Comp Thinking & Data Science

## Lecture 5 – Visualizations

BARNARD COLLEGE OF COLUMBIA UNIVERSITY

# Announcements



- HW01 due tonight (Wednesday 11/04 11:59pm)
- Lab02 ([Data Types and Arrays](#)) due Friday
- HW02 - [Table Manipulation & Visualization:](#)
  - Release tomorrow (Thursday 11/05)
  - Due Monday (11/09)

# Gradescope – Results vs Code View



Results	Code	Results	Code	Results	Code
STUDENT Adam Poliak				STUDENT Adam Poliak	
AUTOGRADE SCORE <b>- / 16.0</b>				AUTOGRADE SCORE <b>- / 16.0</b>	
QUESTION 2				FAILED TESTS	
Question 1	- / 2.0 pts			q2_1 - Public (0.0/1.0) q2_2 - Public (0.0/1.0)	
QUESTION 3				q3_1 - Public (0.0/1.0) q3_2 - Public (0.0/1.0)	
Question 4.3	- / 1.0 pts			q3_3 - Public (0.0/1.0) q4_1 - Public (0.0/1.0)	
QUESTION 4				q4_2 - Public (0.0/1.0) q5_1 - Public (0.0/1.0)	
Question 6.1	- / 2.0 pts			q5_2 - Public (0.0/1.0) q5_3 - Public (0.0/3.0)	
QUESTION 5				q5_4 - Public (0.0/1.0) q8_1 - Public (0.0/1.0)	
Question 6.2	- / 1.0 pts			q7_1 - Public (0.0/1.0)	
QUESTION 6				PASSED TESTS	
Question 6.3	- / 1.0 pts			q9_1 - Public (0.0/0.0)	
QUESTION 7				QUESTION 2	
Question 6.4	- / 1.0 pts			Question 1	- / 2.0 pts
QUESTION 8				QUESTION 3	
Question 9.2	- / 0.0 pts			Question 4.3	- / 1.0 pts
				QUESTION 4	
				Question 6.1	- / 2.0 pts
				QUESTION 5	
				Question 6.2	- / 1.0 pts

# Gradescope – Viewing notebook



Submitted Files for Homework 00

Results

Code

▼ hw01.ipynb Download

```
In [ ]: # Initialize Otter
import otter
grader = otter.Notebook()
```

## Homework 1: Arrays and Tables

```
In [ ]: # Don't change this cell; just run it.
# When you log-in please hit return (not shift + return) after typing in your email
import numpy as np
from datascience import *
```

**Recommended Reading:**

- [Data Types](#)
- [Sequences](#)
- [Tables](#)

Please complete this notebook by filling in the cells provided. Before you begin, execute the following cell to load the provided tests. Each time you start your server, you will need to execute this cell again to load the tests.

This assignment is due Wednesday, November 4th at 11:59 P.M. You will receive an early submission bonus point if you turn in your final submission by Monday, November 2nd at 11:59 P.M. Late work will be accepted up to two days after the assignemnet deadline, as per the [policies](#) of this course.



# Table Review

# Table Review



- `t.sort(column)` sorts rows in increasing order
- `t.sort(column, descending=True)` sorts rows in decreasing order
- `t. take(row_numbers)` keeps the numbered rows
  - Each row has an index, starting at 0
- `t. where(column, are.condition)` keeps all rows for which a column's value satisfies a condition
- `t. where(column, value)` keeps all rows where a column's value equals some particular value
  - Equivalent as `t. where(column, are.equal_to(value))`

# Types of Attributes



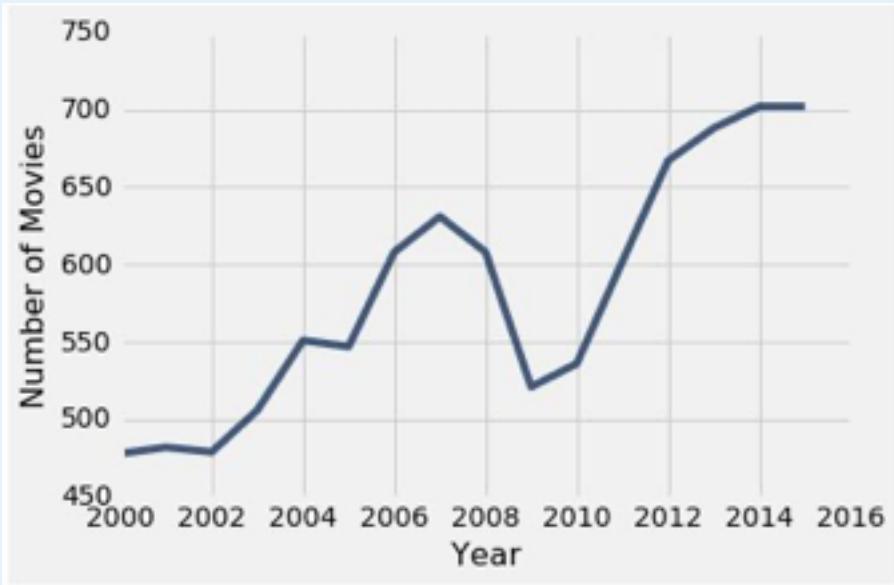
All values in a column of a table should be both the same type and be comparable to each other in some way

- **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 difference

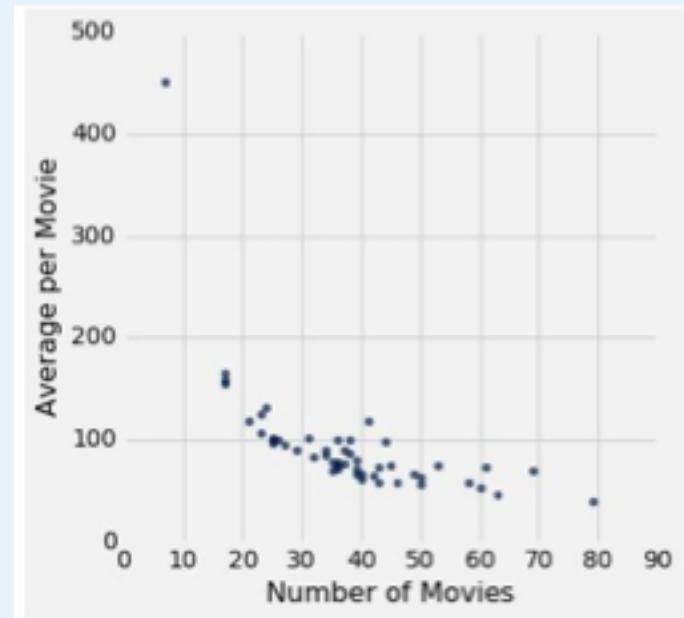
# Plotting Numerical data



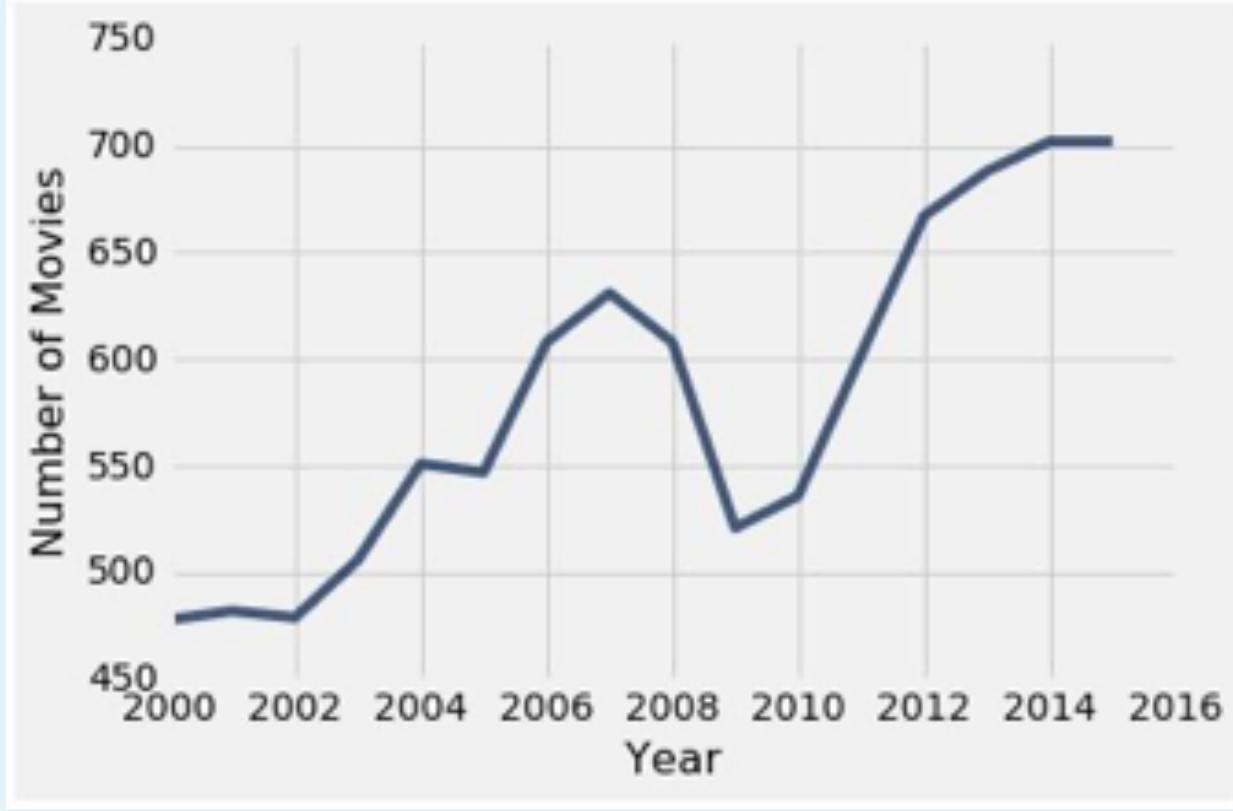
## Line graph plot



## Scatter plot scatter



# x-axis and y-axis



Which is the x-axis?

- Year

Which is the y-axis?

- Number of Movies

# Line vs Scatter plot: When to use which?



- Use **line plots** for sequential data if
  - 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**
  - when looking for associations