

# **CSS 330** Data wrangling and visualization

Instructor: Nabigazinova Elnura

# About the course

Full process of collecting data, preparing data, adapting to specific format and visualizing data



# Data Collection

## Sample sources

- Web sites
- Text files
- Database records
- etc.

## Sample storage formats

- TXT
  - CSV
  - JSON
  - binary
  - etc.
- 

# Data Wrangling

Data wrangling, sometimes referred to as data munging, is the process of transforming and mapping data from one "raw" data form into another format with the intent of making it more appropriate and valuable for a variety of downstream purposes such as analytics.

from [Wikipedia](#)



# Data Analysis

- Statistical analysis
- Machine Learning algorithms
- Natural Language Processing
- Semantic Analysis
- etc.



# Data Visualization

- Reports
- Plots
- Histograms
- Dashboards
- etc.



# Course content

## 10-12 Assignments

These are programming tasks, where you should show your skills

**2 Quizzes** - Progress check

**Midterm Exam** - Midterm project and theory

**Attendance** - Participation is vital to pass the course

**Final Exam** - Final Project and theory



# What we will use

- Python 3
- The Jupyter Notebook
- Unix like OS
- Relational Database (ex: MySQL)
- Dashboarding tools (PowerBI, Tableau)





# Grading Policy

Assignments - 20%

Quizzes - 20%

Midterm - 20%

Final - 30%

Attendance - 10%



# Assignment 0

Install anaconda, and following Python packages:

- numpy
- scipy
- pillow
- matplotlib
- Pandas
- jupyter notebook.

Install untill practice

