

Assignment 2: Data Query and Summarization with Pandas

(60 Points Total, 4 Points for Each Task)

Data available under Resources>Assignment Data.

This exercise focuses on the use of Pandas to query and summarize data. You will need to import the NumPy and Pandas libraries. Produce code to answer the questions below. Deliver your results in a Jupyter Notebook. Make sure all questions and answers are clearly labelled.

The [high_plains_data.csv](#) file contains information for all counties in the high plains region of the United States. The elevation, temperature, and precipitation data were extracted from raster grids provided by the PRISM Climate Group at Oregon State University (<https://prism.oregonstate.edu/>). I also summarized percent forest cover from the 2011 National Land Cover Database (NLCD). NLCD data can be obtained from the Multi-Resolution Land Characteristics Consortium (MRLC) (<https://www.mrlc.gov/>).

Variable	Explanation
NAME_1	County name
STATE_NAME	State in which county occurs
TOTPOP10	Total county population in 2010
per_for	Percent forest cover
elev	Elevation in meters
temp	30-year annual normal mean temperature in Celsius
precip	30-year annual normal precipitation in millimeters

The [portland_park_trees.csv](#) file contains information about individual trees in city parks in Portland, Oregon. These data were obtained from the City of Portland Office of Parks & Recreation (<https://www.portlandoregon.gov/parks/article/433143>).

Variable	Explanation
DBH	Diameter at breast height in inches
Condition	Condition of tree
TreeHeight	Height of tree in feet
CrownWidth	Crown width in feet
Family	Tree family
Genus	Tree genus
Common	Tree common name
Genus_spec	Tree genus and species
Native	Whether or not tree is a native species

Your Tasks

Questions Relating to High Plains Data

- Q1: Which state has the largest number of counties?
- Q2: What was the average county-level percent forest cover for counties in Colorado?
- Q3: Which state has the county with the highest recorded mean elevation?
- Q4: Which state has the largest range of mean temperatures by county?
- Q5: Which state has the smallest total population?
- Q6: How many counties in the dataset have a percent forest cover greater than 30%?
- Q7: What percentage of the counties have a mean elevation greater than 2,500 meters?

Questions Relating to Portland Tree Data

- Q8: How many trees are of the *Quercus* or *Acer* genus?
- Q9: How many trees are of the *Quercus* or *Acer* genus and have a DBH larger than 50 inches?
- Q10: Which genus has the highest mean DBH of the following genera: *Quercus*, *Acer*, or *Fraxinus*?
- Q11: How many different species of trees are recorded in the *Acer* genus?
- Q12: How many trees are of the *Acer* genus but not of the *Acer saccharinum* species and have a DBH between 30 inches and 60 inches?
- Q13: What percentage of trees in the *Acer* genus has a condition of Fair?
- Q14: What is the average height of all the Douglas-fir trees in the dataset?
- Q15: What percentage of all the trees in the dataset are native but not of the *Quercus* or *Acer* genus?

Deliverables

- Jupyter Notebook