

Assignment 10: Data Query and Summarization with Pandas

(60 Points Total, 6 Points for Each Task)

Data available under Resources>Portland Data.

The `portland_data.gpkg` GeoPackage contains three data layers for Portland, Oregon:

neighborhoods: boundaries of Portland neighborhoods

park_trees: mapped trees in parks

street_trees: mapped trees along streets

The **neighborhood** boundaries were obtained from the City of Portland (<https://www.portlandoregon.gov/28130>) while the **tree** inventory data were obtained from the City of Portland Office of Parks & Recreation (<https://www.portlandoregon.gov/parks/article/433143>).

To complete this exercise, you will need to have set up a **conda environment** with the following: (1) **NumPy**, (2) **Pandas**, (3) **GeoPandas**, and (4) all dependencies for **GeoPandas**. In this exercise, you will only work with the `park_trees` dataset. Also, you do not need the spatial information. The spatial information can be removed using the following code:

```
trees = gpd.read_file("YOUR FOLDER PATH/portland_data.gpkg", layer='park_trees')
trees2 = pd.DataFrame(trees.drop(columns='geometry'))
```

This exercise will focus on the use of Pandas to query and summarize data.

Your Task

Produce code to complete the requested tasks. Deliver code in a Python or plain text file.

Task 1: Import all required libraries, read in the `park_trees` layer from the `portland_data.gpkg` GeoPackage as a GeoPandas DataFrame, and remove the geometry information.

Task 2: Subset out just the trees in the *Quercus* or *Acer* genera (Hint: use the "Genus" attribute).

Task 3: Subset out just the trees in the *Quercus* or *Acer* genera that have a DBH larger than 50 (Hint: use the "Genus" and "DBH" attributes).

Task 4: Generate a table of mean DBH by genera.

Task 5: Obtain a list of all unique genus names in the data set.

Task 6: Obtain a list of all unique species names in the data set for just the *Acer* genus (Hint: use the "Genus" and "Genus_spec" attributes).

Task 7: Count the number of trees of each species in the *Acer* genus.

Task 8: Obtain the mean DBH for each species in the *Acer* genus.

Task 9: Extract all trees in the *Acer* genus that are of the species *Acer saccharinum* and have a DBH between 30 and 60.

Task 10: Obtain a variety of summary statistics with the `describe()` method for tree height by genera (Hint: use the "TreeHeight" attribute).

Deliverables

- Code in Python or plain text file.