Instructions

Today we're trying to get comfortable with `pandas`, a fantastically useful (but complicated) library for data science. We'll frame this as a scavenger hunt. We'll learn pandas by figuring out how to do successively more complicated things, and we'll do it hands-on.

Your Tasks

• Make a new ipynb notebook in the developer environment of your choice. I know some of you are using Jupyter notebooks locally, while some of you use Google colab. I personally like Google colab or Visual Studio Code. Talk to me if this step seems difficult. • For each question create a new cell that demos how to do each thing in pandas. • Work in small groups • Try to work in order - the questions build on each other.

Questions

Download the Mt. Washington weather data csv file from Canvas. (You can find it under Files->Mt_Washington.csv) Then figure out how to do each of the following:

1. Import the package `pandas`.
2. Load the csv file into a pandas dataframe.
3. Print a list of all the column headings.
    - For each column heading, identify if the variable is a numeric variable or a categorical variable.
        - If it is numeric, identify if the variable is continuous or discrete.
        - If it is categorical, identify if the variable is binary or not.
4. Print out how many rows are in the data.
5. Show what the first few rows look like.
6. Print out row 245.
7. Print out just one column (your choice).
8. Count how many days had some precipitation, and how many had some snow.
9. What's the difference between precipitation and snow? Count how many days had snow but no precipitation, and how many had precipitation but no snow.
10. Simplify things. Drop all of the columns except for these: DATE, PRCP, SNOW, TMIN, TMAX.

Thank you Kyle Wilson, for the original lesson plan, which I have modified.