

# Creating an R Workflow: Case Study of Hop Growth

**GEOG/ECOL/ENR 5050**

**Fall 2016**

**Instructor: Dr. Shannon E. Albeke**

# Objective: Create a workflow to look at climate effect on hop yield

- Data are all 'cloud based'
  - //klondike/shannon\$/TDM/ClassShare/HopAnalysis
- Data include:
  - Hop yield by State/Region and years (tab-delimited; HopYield\_Acre.txt)
  - Shapefile of hop growing regions
  - PRISM climate data for the entire US (2005-2013)
    - Max Temp, annual and by month
    - Precipitation, annual and by month
- I need a regression analyses of annual (i.e. each year) hop yield given annual maxtemp and ppt
  - $YPA \sim \text{Annual\_maxtemp} + \text{Annual\_ppt}$
  - $YPA \sim \text{mean\_maxtemp}[\text{months 4-9}] + \text{mean\_ppt}[\text{months 4-9}]$

# Workflow Outline

1. Create a data.frame of your PRISM datasets
  - Columns: SampYear, SampMonth, maxt\_path, maxt\_file, ppt\_path, ppt\_file
  - I do this so I can more easily reference attributes of files stored within a directory because often time we store metadata in the file name
2. Bring in you hop yield dataset and reformat to match how a regression analysis needs to be (i.e. think about each row of data as an observation in space and time, what do you need to match the regression formula:  $YPA \sim \text{maxtemp} + \text{ppt}$ )

# Workflow Outline, cont'd

3. Read in your shapefile of hop regions
4. Loop through your raster data.frame and 'stack' the rasters representing annual maxtemp and ppt
  - You may want to deal the 'names' of your rasters to make them easier to understand in future steps (multiple ways to do this, personal preference)
5. After stacking the rasters, extract the mean values for each polygon
6. Format the mean raster value data.frame to match your hop yield and then merge the two tables together
7. Fit a linear regression
8. Repeat this process, but use the specific months 4-9 (taking the mean values of these months, thus one extra raster analysis step using the raster::calc function) for your rasters instead of the annual (i.e. create your own annual rasters from a more limited range of months)