**Reproducible project set-up**

Reproducibility is a cornerstone of good scientific research and in this exercise you will design a workflow and make your own ‘project’ template to improve reproducibility of your work.

This will be done in two parts.

1) design a workflow in R.

2) apply this to an actual project.

***1) Design a workflow in R***

*15 minutes break-out*

With your group members come with a general workflow for a project. Use our readings for guidance.

Start to plan how to implement this as an RStudio project set-up (different R scripts for different parts of your workflow).

Keep this part schematic.

In class we will discuss the different approaches and derive a framework and work on setting his up in R.

***2) Apply your workflow to an actual project: Canadian weather data.***

You can access weather station data online using the "weathercan" package. Check out their GitHub for a basic crash course in how to access the data: <https://github.com/ropensci/weathercan>

This project is part of ROpenSci (<https://ropensci.org/>) which has lots of great projects like this.

The task is to put your own work flow to use to try to find out how well the maximum temperature and precipitation varies from west to east across the Coast Mountains. Mountain ranges strongly influence rainfall but how does it affect max temperature? Is the pattern the same?

To do this, select a series of weather stations, starting at Vancouver, next is Squamish, Pemberton, Lillooet and you can get more stations if you like.

a) Figure out how to get weather data (one entry per day) for the different stations for a one-year period. Download this data for the stations of interest.

b) Visualise the change through of max temp and precipitation for the year. What general patterns can you see? (Remember: the graph tells a story!). Save the plots as pdf.

c) Calculate the correlation coefficients for all pairwise comparisons between stations possible (Vancouver – Squamish, Vancouver – Pemberton, ..) for max temp and precipitation. Save the tables of coefficients as csv files.

How do the correlation coefficients change with geographical distance between stations? (You can use e.g. Google earth to get these distances (as the-crow-flies) between stations; approximate this.). Think of a good way to visualise this and save plots as pdf.

Do these patterns differ for max temperature and precipitation? Why do you think?

Make sure that all the code is well annotated with comments.

**New functions of interest**

paste()

dir.create()

file.exists()

list.files()

pdf() ‘*graphics code*’ dev.off()

cor()