DSCI 100 - Introduction to Data Science

Lecture 1 - Getting started with Jupyter & R

2019-01-03

High-level goals of this course:

1. Learn how to use reproducible tools (Jupyter + R) to do data analysis

1. Learn how to solve 3 common problems in Data Science

Problems we will focus on:

1. Predict a class/category for a new observation/measurement (*e.g.*, cancerous or benign tumour)

1. Find previously unknown/unlabelled subgroups in your data (*e.g.*, products commonly bought together on Amazon)

1. Predict a value for a new observation/measurement (*e.g.*, 10 km race time for 30-35 year old males with a BMI > 25).

Course syllabus:

Read on your own time: https://github.com/UBC-DSCI/dsci-100/blob/master/README.md (https://github.com/UBC-DSCI/dsci-100/blob/master/README.md)

TL;DR

Well, please do read the syllabus later... but for now...

Flipped classroom

- read text/watch videos before class
- lecture worksheets and activities in class (Thursdays), due Saturdays at 6pm
- tutorial questions in class (Tuesdays), due Wednesdays at 10pm
- you will need a laptop/chromebook/etc in every class! Don't have one? Borrow one from the library (see here (https://services.library.ubc.ca/computers-technology/technology-borrowing/)).

Everything will be posted as links/buttons in Canvas (https://canvas.ubc.ca/courses/19078)

Collaborate

- talk to each other (in class, on Piazza) as you work through the worksheets and tutorials
- group project at middle-end of course

Your teaching team

Position	Name
Instructor	Tiffany Timbers
Teaching Assistant	Madison Friesen
Teaching Assistant	Harmeet Gill
Teaching Assistant	Aaron Quinton

First week learning goals:

- use a Jupyter notebook to execute provided R code
- edit code and markdown cells in a Jupyter notebook
- create new code and markdown cells in a Jupyter notebook
- load the `tidyverse` library into R

- create new variables and objects in R using the assignment symbol
- use the help and documentation tools in R
- match the names of the following functions from the `tidyverse` library to their documentation descriptions: `read_csv`, `select`, `mutate`, `filter`, `ggplot`, `aes`
- chain together two functions using the pipe operator, `%>%`

We've got a lot to do! Let's get started!

- Everyone, navigate to Canvas (https://canvas.ubc.ca/courses/19078) and open the assignment `worksheet_01` (https://canvas.ubc.ca/courses/19078/assignments/281853).

- Use your neighbours, the TAs and me to help you get unstuck when needed!

- I will interupt in about 20 minutes for a class activity.