

# Welcome to data science Friday!

Kim Dill-McFarland – kadm@mail.ubc.ca

Steven Hallam – shallam@mail.ubc.ca



# Learning objectives

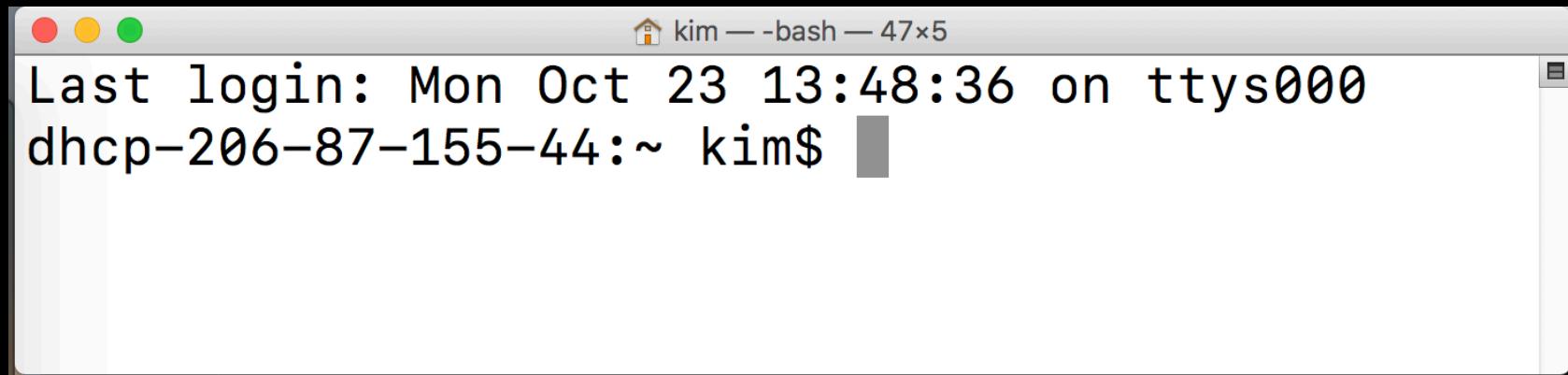
Students will be able to

- Identify the strengths and potential uses of command line and RStudio
- Create appropriate file names and directory structures

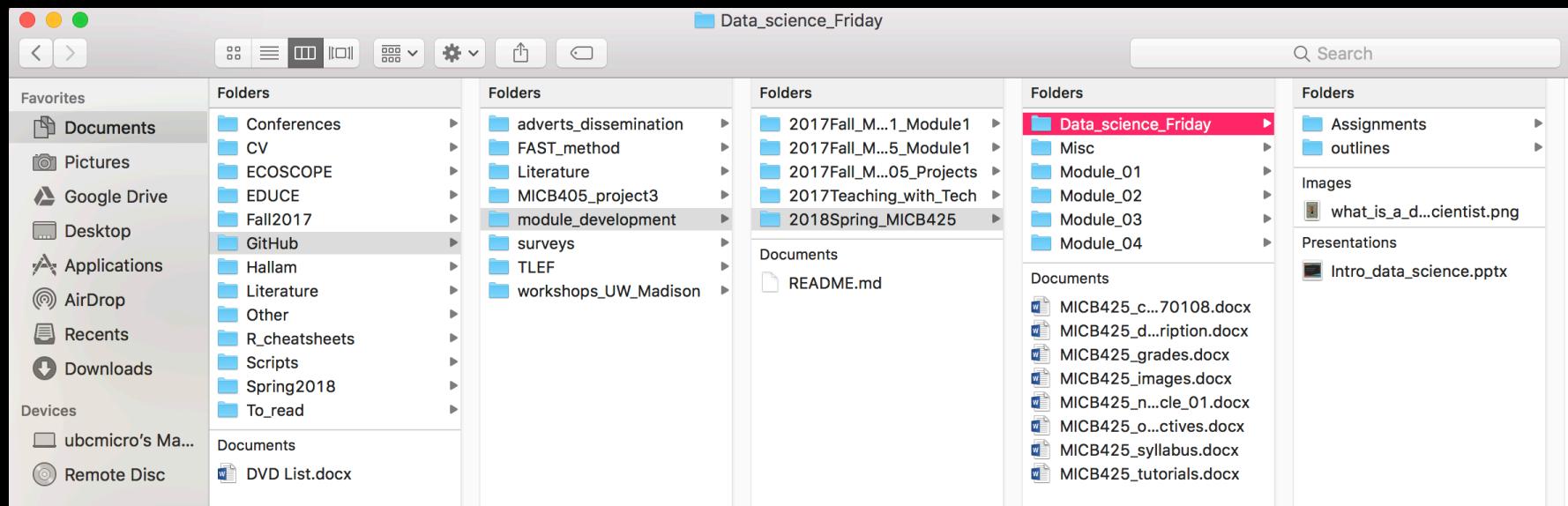
# Tools in this course

- Command line terminal
- RStudio
- GitHub
- Metagenomics tools

# Command line terminal



```
kim — -bash — 47x5
Last login: Mon Oct 23 13:48:36 on ttys000
dhcp-206-87-155-44:~ kim$
```



# Command line can...

- Do everything you do in Finder/File explorer
  - Make, move, copy, delete, rename files and folders
- Run specific programs
  - Text editors, R, sequence analysis tools
- Connect to and transfer files from remote servers like GitHub
- And more...

# Why command line?

- There are processes that are easier, faster, or can only be accomplished using the command line
  - Super computing resources can only be accessed via command line
- Command line allows you to provide reproducible code for your future self as well as supervisors and co-workers
- Commands are the same across different OS types and versions
- Command line knowledge makes you more competitive not only for careers in biology but also in other STEM areas, finance, etc...

# RStudio

~/Documents/Fall2017/Micro301/MICB301\_Fall2017 - RStudio

MICB301\_Fall2017

Console Terminal ~ /Documents/Fall2017/Micro301/MICB301\_Fall2017/

```
R version 3.4.2 (2017-09-28) -- "Short Summer"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

> |

Environment History Connections

Import Dataset List Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Fall2017 > Micro301 > MICB301\_Fall2017

Name	Size
..	
.Rhistory	67 B
Images	
MICB301_Fall2017.html	5.9 M
MICB301_Fall2017.Rmd	21.4
MICB301_Fall2017.Rproj	205
Module1_OTU.fasta	8 ME
Module1_OTU.txt	1.8 M
Module1_taxonomy.txt	4.6 M

# RStudio can...

- Data manipulation
- Graphing
- Statistics

# Why RStudio?

- Open source
- Huge community of developers around the world
- Custom packages
- Reproducible research with Rpubs and Rmarkdown

<http://rpubs.com/dillmcfarlan>



Using command line-esque  
programs...

Stay organized!!

# File naming

Of the following, which is the best file name and why?

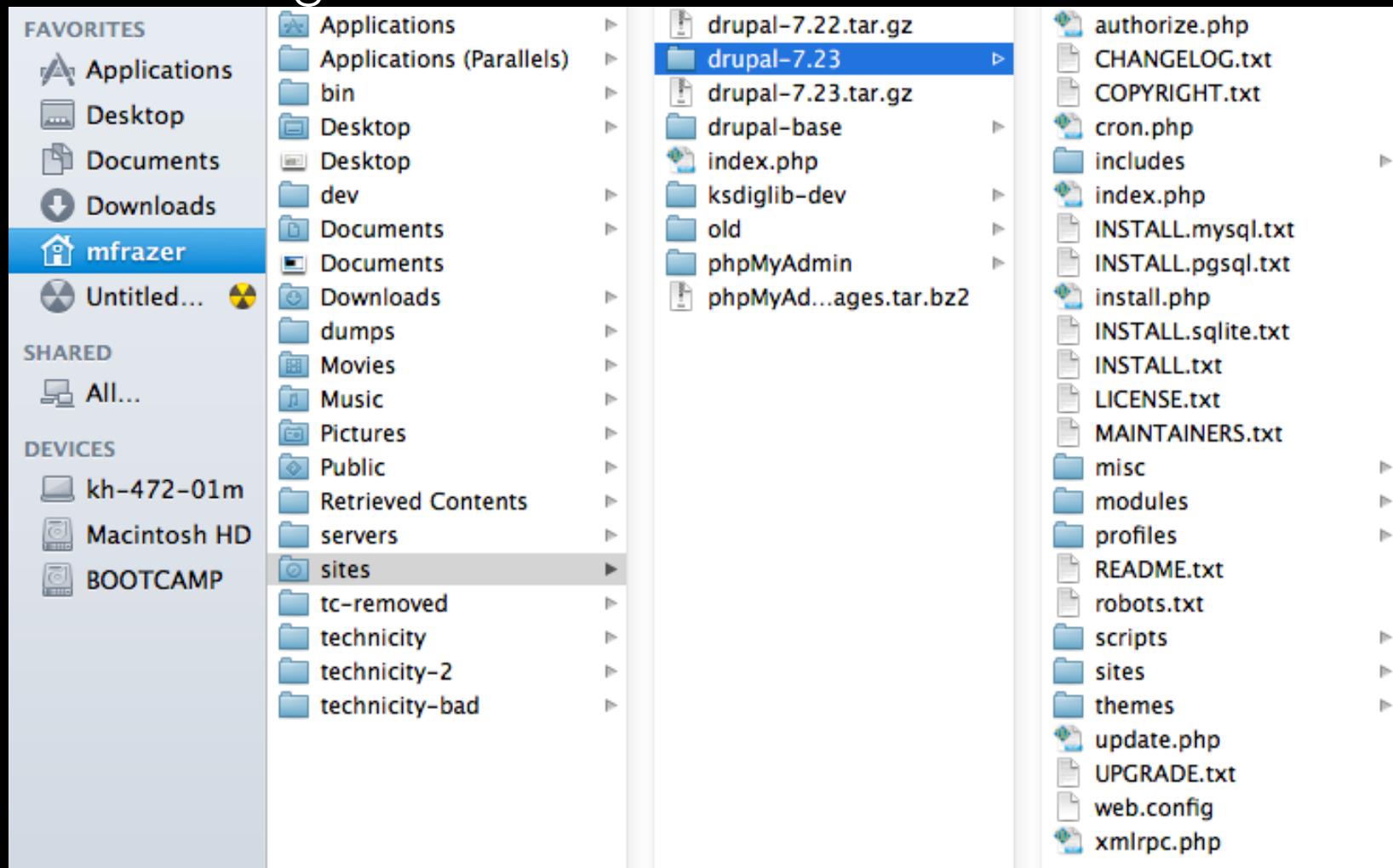
- A. Portfolio.txt
- B. 2018\*Portfolio(MICB425).txt
- C. 2018\_MICB425\_portfolio\_DillMcFarland.txt
- D. DillMcFarland\_porfolio\_MICB425\_2018.txt
- E. Portfolio MICB425 2018.txt
- F. 2018\_term2\_MICB425\_portfolio\_Kim\_DillMcFarland\_ID12345678.txt

# Guidelines for file names

- Make it unique
- Include enough information to know what the file is without being extremely long
- Avoid special characters especially \*, (), and “”
- Only use capital letters when they have meaning
- Use \_ in between words and leave . for just before the file extension
- Order pieces to result in desired alphabetical sorting

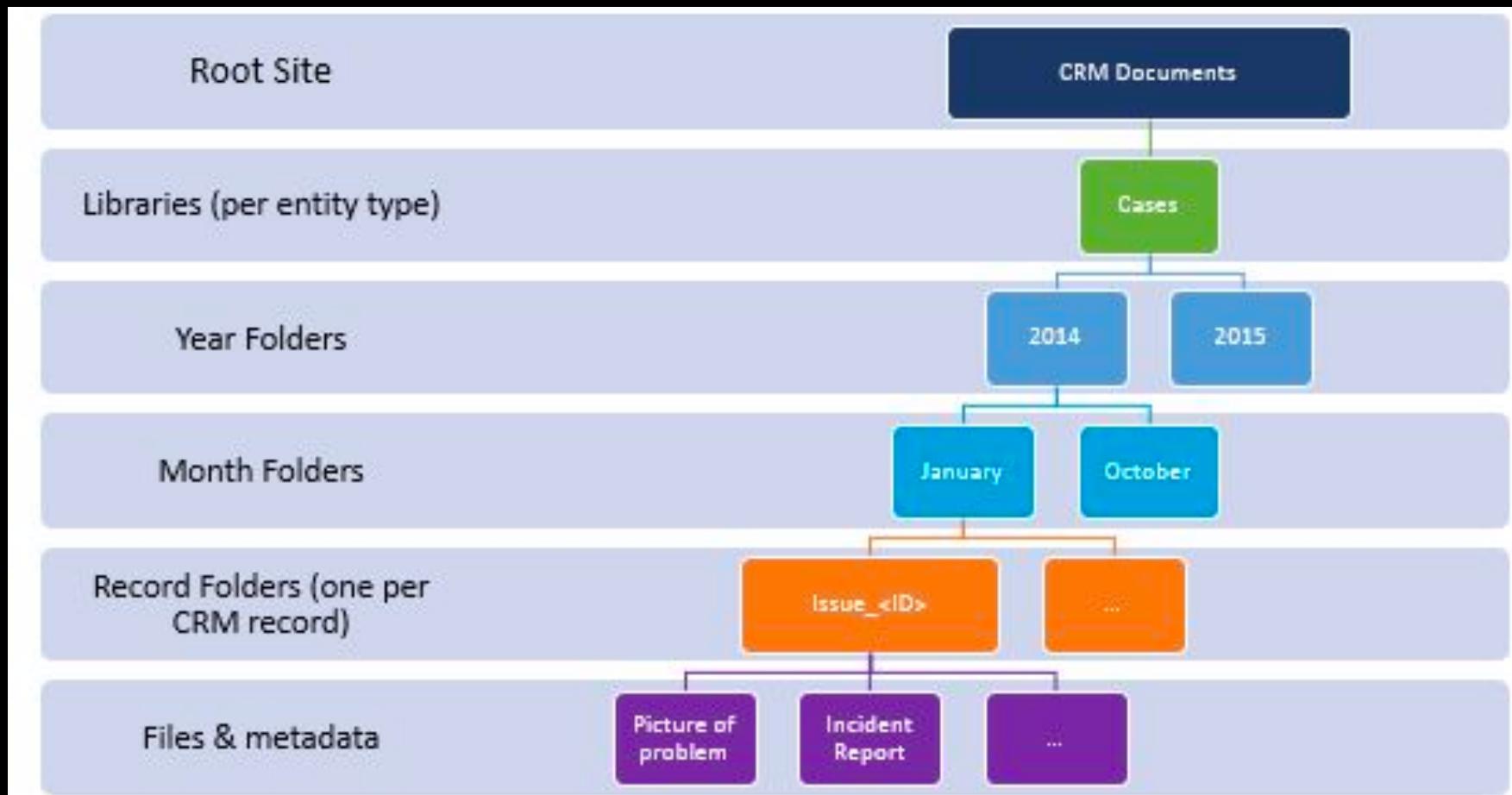
# Directory structure

What's wrong with this structure?



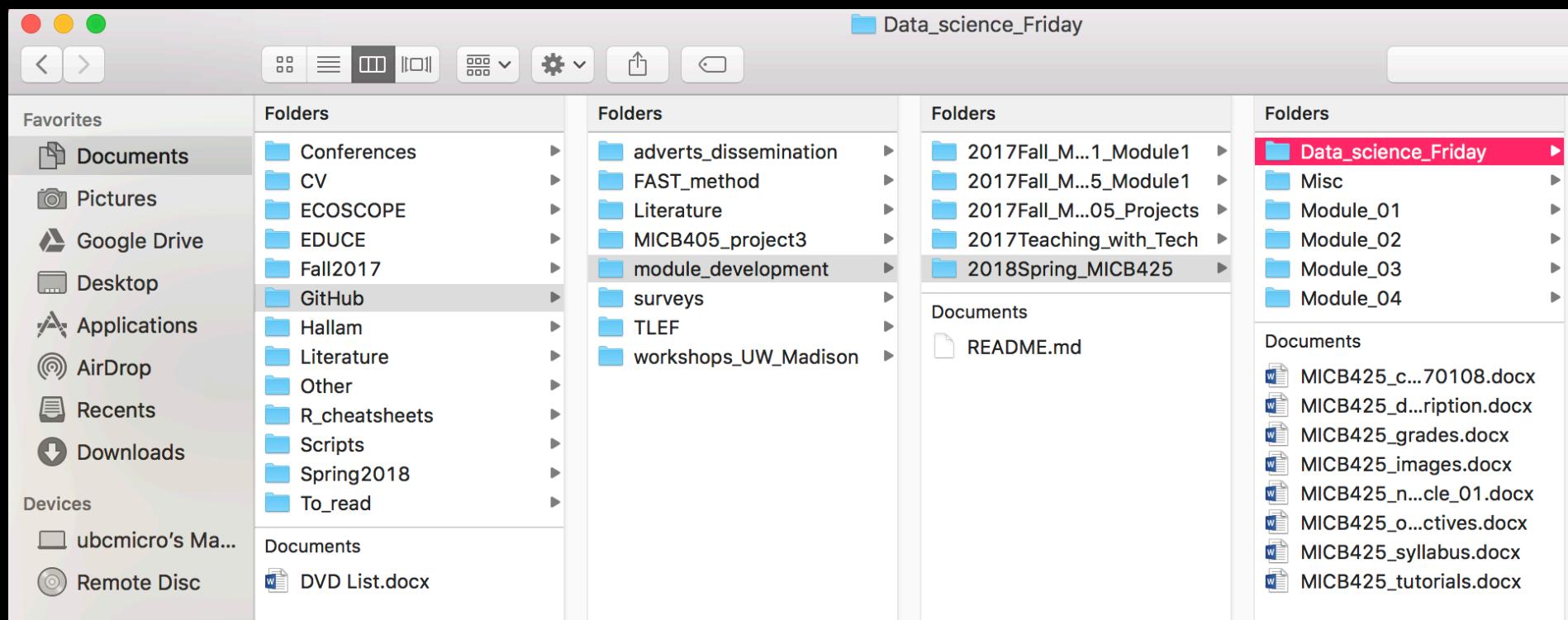
# Directory structure

What's wrong with this structure?



# Directory structure

What's wrong with this structure?



# Guidelines for file structure

- Make it unique. Think of directory names as key words
- Avoid special characters especially \*, (), and “”
- Only use capital letters when they have meaning
- Use \_ in between words
- Order pieces to result in desired alphabetical sorting
- Consider a template structure for similar folders like for each of your courses

# GitHub

- Code hosting, review, sharing, editing
- Project and team management
- Version control

# Naming versions

"FINAL".doc



FINAL.doc!



FINAL\_rev.2.doc



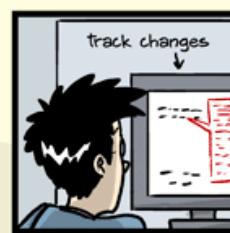
FINAL\_rev.2.doc



FINAL\_rev.6.COMMENTS.doc



FINAL\_rev.8.comments5.  
CORRECTIONS.doc



FINAL\_rev.18.comments7.  
corrections9.MORE.30.doc



FINAL\_rev.22.comments49.  
corrections.10.#@\$%WHYDID  
ICOMETOGRAD SCHOOL?????.doc





# GitHub

<https://github.com/EDUCE-UBC/MICB425>

# Data science workshops

- R/RStudio
- Python
- GitHub and reproducible research

<http://ecoscope.ubc.ca/events/>

# DS Assignment 1

Due *before class* Friday, Jan 12

- Download and install
  - Git Bash command line (Windows only)
  - R and RStudio
- Create GitHub account
- Office hours
  - Wed Jan 10 @ 10-11AM in LSI 1510
  - Wed Jan 10 @ 2-3PM in LSI 1416
  - Thurs Jan 11 @ 2-3PM in LSI 1416

# Survey consent options

Option	Complete surveys for credit	Include responses in EDUCE study	Complete alternate assignment for credit
A	X	X	
B	X		
C			X