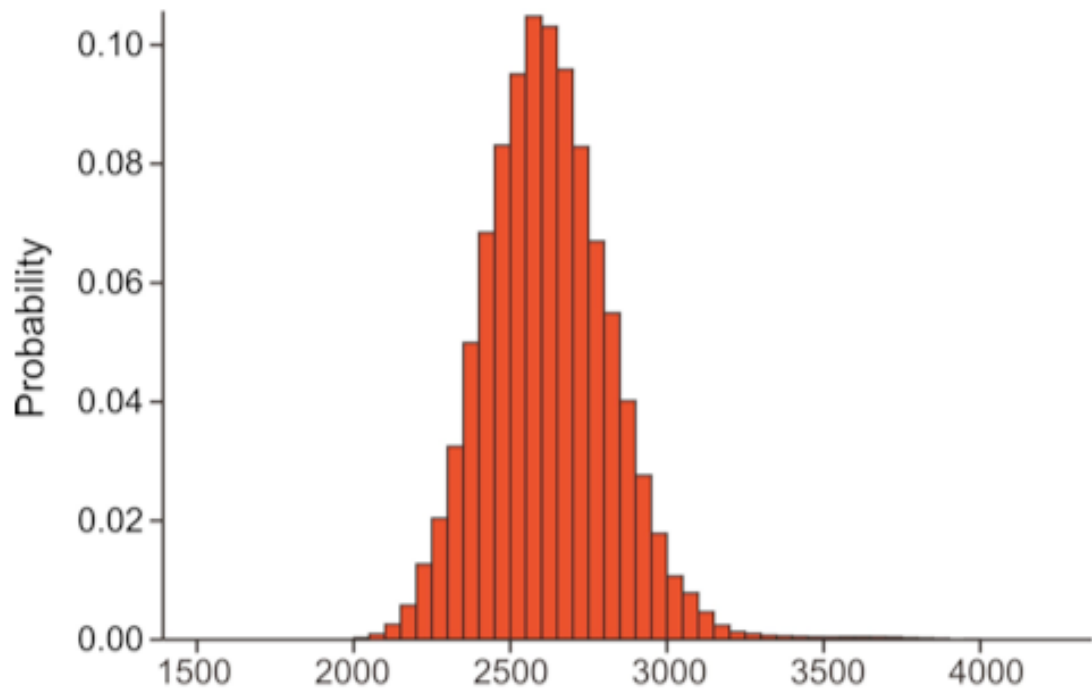
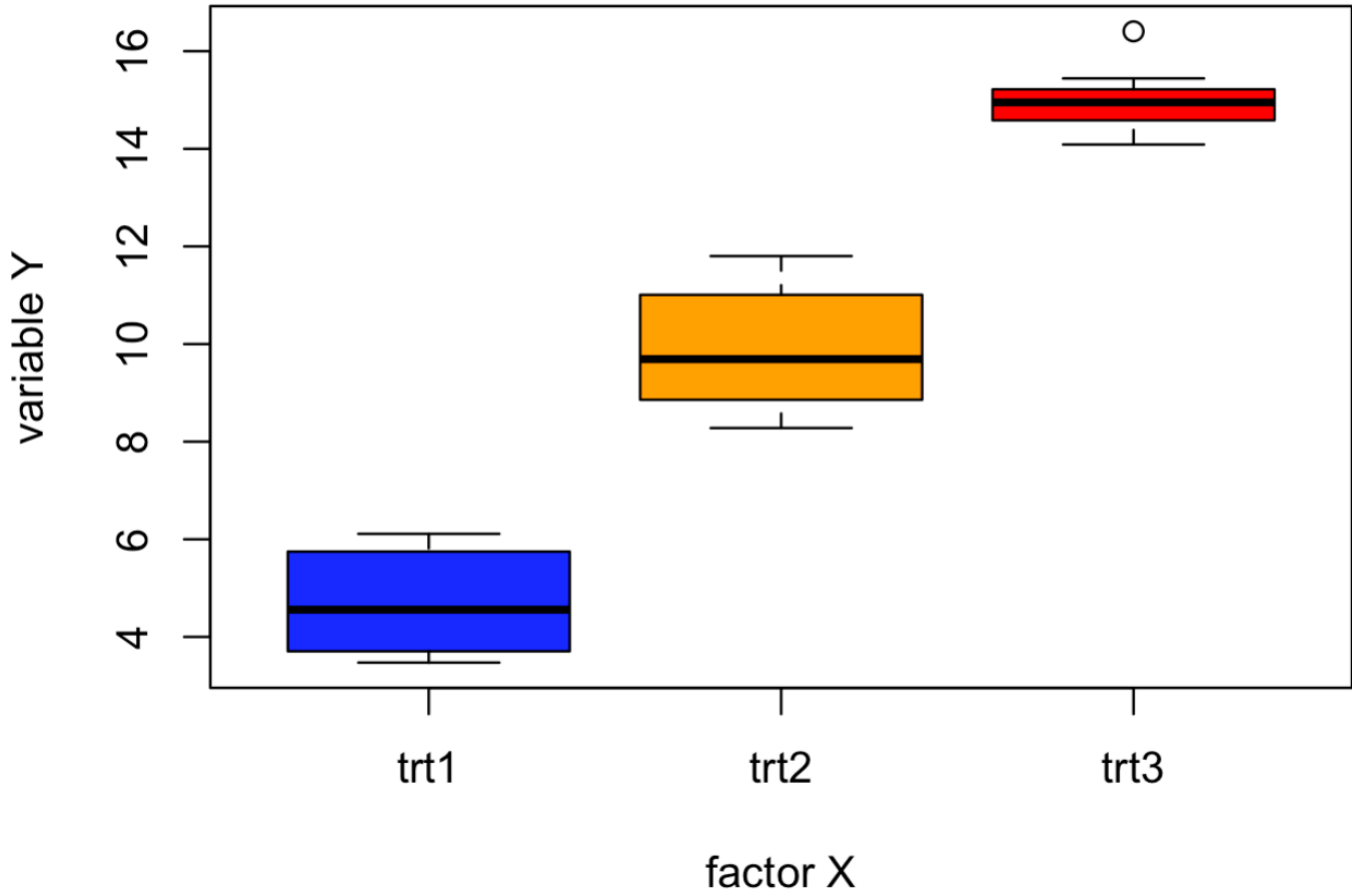


# Foundational Statistics for Data Science

## DSCI 626



`boxplot(y~x)`



A diagram showing a 5x10 grid of pots. Each pot contains a green plant. The pots are colored red and blue in a checkerboard pattern, alternating by row and column.									
One Measurement Per Pot					Three Measurements Per Pot				
Total Sample Size (N) = 50					Total Sample Size (N) = 150				
Experimental Units = 50					Experimental Units = 50				
Treatments = 2					Treatments = 2				
Treatment Size (n)= 25					Treatment Size (n)= 25				
Replication (r) = 25					Replication (r) = 25				

# Your Instructor

**Clay Small (he/him/his), PhD**

- BS in Zoology from the University of Idaho
- PhD in Zoology from Texas A&M University
- Postdoc at UO (Cresko Lab in IE<sup>2</sup>)
- Currently a Research Assistant Professor in the DS Department

# **Student Introductions!**

**Let's get to know each other a bit**

- Your preferred first name, your last name, and your pronouns
- Your year, department/institute, and area of interest
- One thing that you would like to take away from this course

# Course Overview

- A **true introduction** to statistical thinking and doing (R)
- Active learning! (Minimal lecture and lots of time for in-class work)
- Collaboration! (Let's try to emphasize group work, in and outside of class)
- A safe space for **all** questions

**Let's take a look at the syllabus/schedule**

# Primary Goals and Learning Outcomes

## Learn fundamental skills for your careers in science

- Basic computational experience (Unix and R/RStudio)
- Understand core statistical concepts for application
- Learn effective experimental design and data analysis
- Prepare for more advanced stats courses (e.g. “Advanced Biological Statistics” in the Fall)

# Primary Goals and Learning Outcomes

## Some specific knowledge

- Understand and write your own code in R (and Unix)
- Understand probability distributions and sampling
- Understand p-values, test statistics, and types of statistical error
- Learn to use statistical analyses like t-tests, ANOVA, non-parametric tests, and frequency-based tests
- Make plots for figures (using R) that are publication-quality
- Learn some basics of sound experimental design
- Learn some of the history (negative and positive) of statistics

# Why do we need to write code?

- We often need to repeat operations hundreds or thousands of times
- Manipulating/reformatting large text files using GUI software is slow and error-prone
- Allows us to implement powerful open-source tools written by others
- Cross-platform utility (Mac, Windows, Linux, HPCs, etc.)
- Any other reasons that y'all think are important?

# Why do we need statistics?

- We almost never know the world perfectly - need to make predictions and decisions, though
- We need to estimate parameters based on samples of data
- Science depends heavily on testing hypotheses
- Complex/Large data sets can't be easily visualized in their own right
- Math and computers facilitate all of the above



# Working from a “terminal”

For the first few tasks (today and Thurs.), we’ll work on managing text files using Unix commands

```
ubuntu@ip-10-4-230-31: ~/working — ssh — 141x44
ubuntu@ip-10-4-230-31:~/working$ ls -la ~/
total 444
drwxr-xr-x 18 ubuntu ubuntu 4096 2012-01-09 22:50 .
drwxr-xr-x  6 root   root   4096 2011-11-14 17:12 ..
-rw-r----- 1 ubuntu ubuntu 3757 2012-03-12 12:11 .bash_history
-rw-r--r--  1 ubuntu ubuntu  220 2011-05-18 10:00 .bash_logout
-rw-r--r--  1 ubuntu ubuntu 3581 2011-11-14 17:16 .bashrc
lrwxrwxrwx  1 root   root    21 2011-11-14 12:25 bin -> ../../usr/proftpd/bin
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 17:16 .byobu
drwxrwxr-x  4 ubuntu ubuntu 4096 2011-11-14 14:33 .cabal
drwx----- 3 ubuntu ubuntu 4096 2011-11-14 11:37 .cache
lrwxrwxrwx  1 root   root    20 2011-11-14 12:23 conf -> ../../usr/nginx/conf
-rwxrwxrwx  1 ubuntu ubuntu  992 2011-11-14 17:12 configure_freemx.sh
drwx----- 3 ubuntu ubuntu 4096 2012-01-08 22:52 .emacs.d
lrwxrwxrwx  1 root   root    21 2011-11-14 12:25 etc -> ../../usr/proftpd/etc
drwxr-xr-x  2 ubuntu ubuntu 4096 2011-11-14 12:51 .fontconfig
drwx----- 2 ubuntu ubuntu 4096 2011-11-14 14:18 .gconf
drwxr-xr-x  3 root   root   4096 2011-11-14 16:58 .gem
drwx----- 2 ubuntu ubuntu 4096 2011-11-14 14:51 .gnupg
lrwxrwxrwx  1 root   root    20 2011-11-14 12:23 html -> ../../usr/nginx/html
lrwxrwxrwx  1 root   root    25 2011-11-14 12:25 include -> ../../usr/proftpd/include
drwxrwxr-x  4 ubuntu ubuntu 4096 2011-11-28 17:49 install
drwxrwxr-x  3 ubuntu ubuntu 4096 2011-11-14 12:27 .lein
-rw-r----- 1 ubuntu ubuntu  65 2011-11-14 13:07 .lessht
lrwxrwxrwx  1 root   root    21 2011-11-14 12:25 lib -> ../../usr/proftpd/lib
lrwxrwxrwx  1 root   root    25 2011-11-14 12:25 libexec -> ../../usr/proftpd/libexec
lrwxrwxrwx  1 root   root    20 2011-11-14 12:23 logs -> ../../usr/nginx/logs
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 12:51 .m2
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 14:18 .matplotlib
-rw-r----- 1 ubuntu ubuntu 2964 2012-01-09 22:50 .mysql_history
-rw-r--r--  1 ubuntu ubuntu  675 2011-11-14 17:16 .profile
drwxr-xr-x  2 root   root   4096 2011-11-14 12:25 sbin
-rw-rw-r--  1 ubuntu ubuntu  0 2011-11-14 11:37 .screenrc
lrwxrwxrwx  1 root   root    23 2011-11-14 12:25 share -> ../../usr/proftpd/share
drwx----- 2 ubuntu ubuntu 4096 2011-11-14 11:33 .ssh
-rw-rw-r--  1 ubuntu ubuntu 338416 2012-01-09 16:12 stacks-0.998.tar.gz
drwxrwxr-x  3 ubuntu ubuntu 4096 2011-11-14 13:44 .subversion
-rw-r--r--  1 ubuntu ubuntu  0 2011-11-14 11:38 .sudo_as_admin_successful
drwxrwxr-x  2 ubuntu ubuntu 4096 2012-01-09 04:42 tmp
lrwxrwxrwx  1 root   root    21 2011-11-14 12:25 var -> ../../usr/proftpd/var
-rw-r----- 1 ubuntu ubuntu 7522 2012-01-09 20:35 .viminfo
lrwxrwxrwx  1 ubuntu ubuntu  12 2012-01-08 22:49 working -> /mnt/working
-rw-r----- 1 ubuntu ubuntu  196 2011-12-12 12:04 .Xauthority
ubuntu@ip-10-4-230-31:~/working$
```

# For the remainder of class...

- If you don't have a Unix/Linux Terminal (i.e. your Windows OS doesn't have an app pre-installed), install the Windows Subsystem for Linux

(Here is one guide: <https://canonical-ubuntu-wsl.readthedocs-hosted.com/en/latest/guides/install-ubuntu-wsl2/>)

- **As a class** - We will practice navigating from the terminal, making new directories, using various commands like `man`, `cat`, `paste`, `grep`, etc.
- **On your own** - Check out Chapter 5 in the course Gitbook

# UO Libraries Data Services Help Desk

<https://library.uoregon.edu/departments-of-open-research/data-services/data-services-consultations>

- Where? Monday-Friday, 11am - 4pm, in the Knight Library
- Also offer workshops on R, Python, Git, etc.
- Check out their website