Stat 411/511

WELCOME

Sep 25th 2015

Today

Introductions
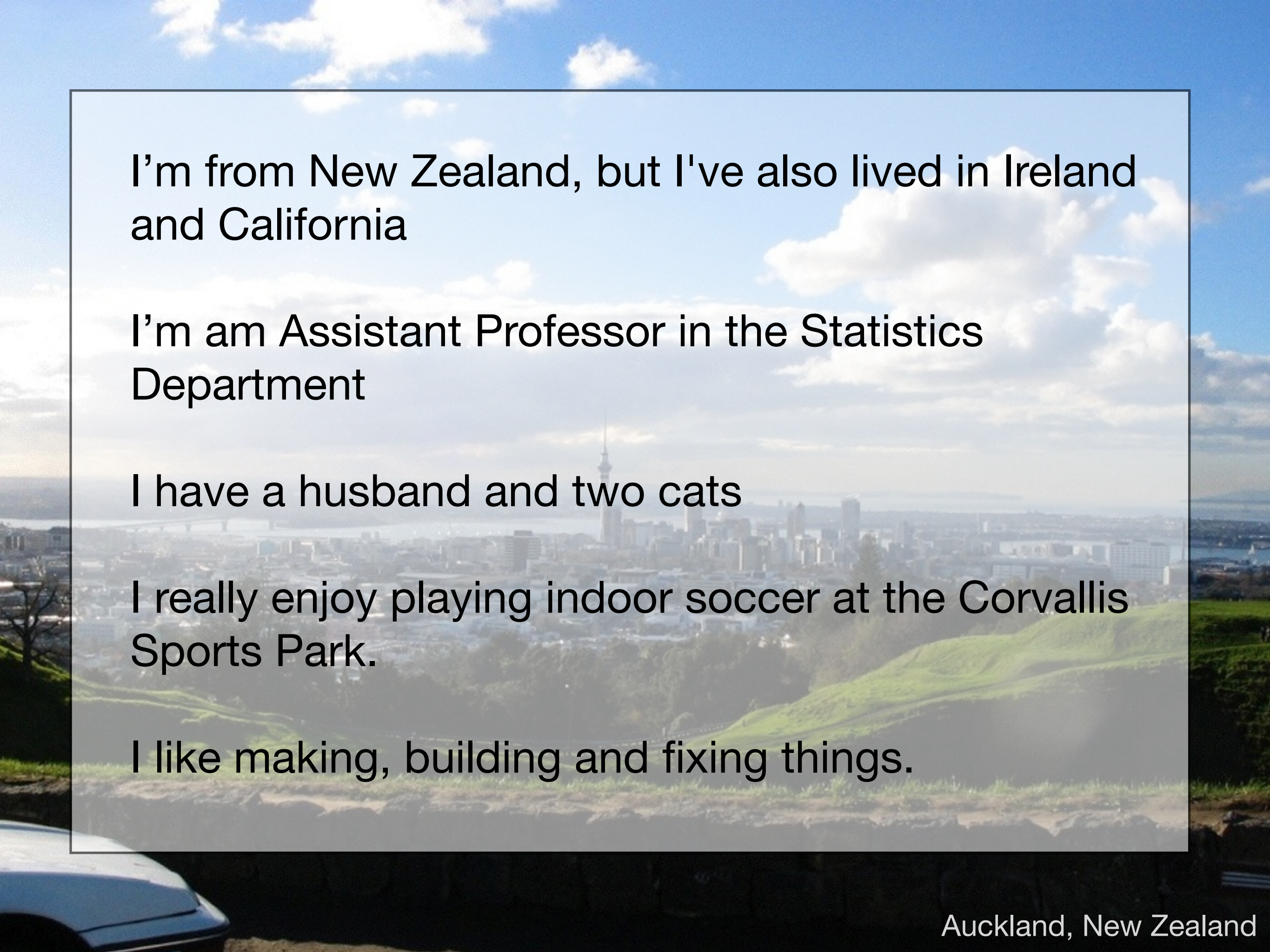
What do we learn in ST411/511?

Syllabus

R & RStudio

Hello, my name is

Charlotte

Office: Weniger 255 (soon?)
Email: wickhamc@stat.oregonstate.edu

I'm from New Zealand, but I've also lived in Ireland and California

I'm am Assistant Professor in the Statistics Department

I have a husband and two cats

I really enjoy playing indoor soccer at the Corvallis Sports Park.

I like making, building and fixing things.

Auckland, New Zealand

Your turn

Introduce yourself to your neighbour
Why are you taking ST411/511?

Four important people who aren't here, your TAs:
Ben, Shaoshan, Spencer and Chris

stat511.cwick.co.nz
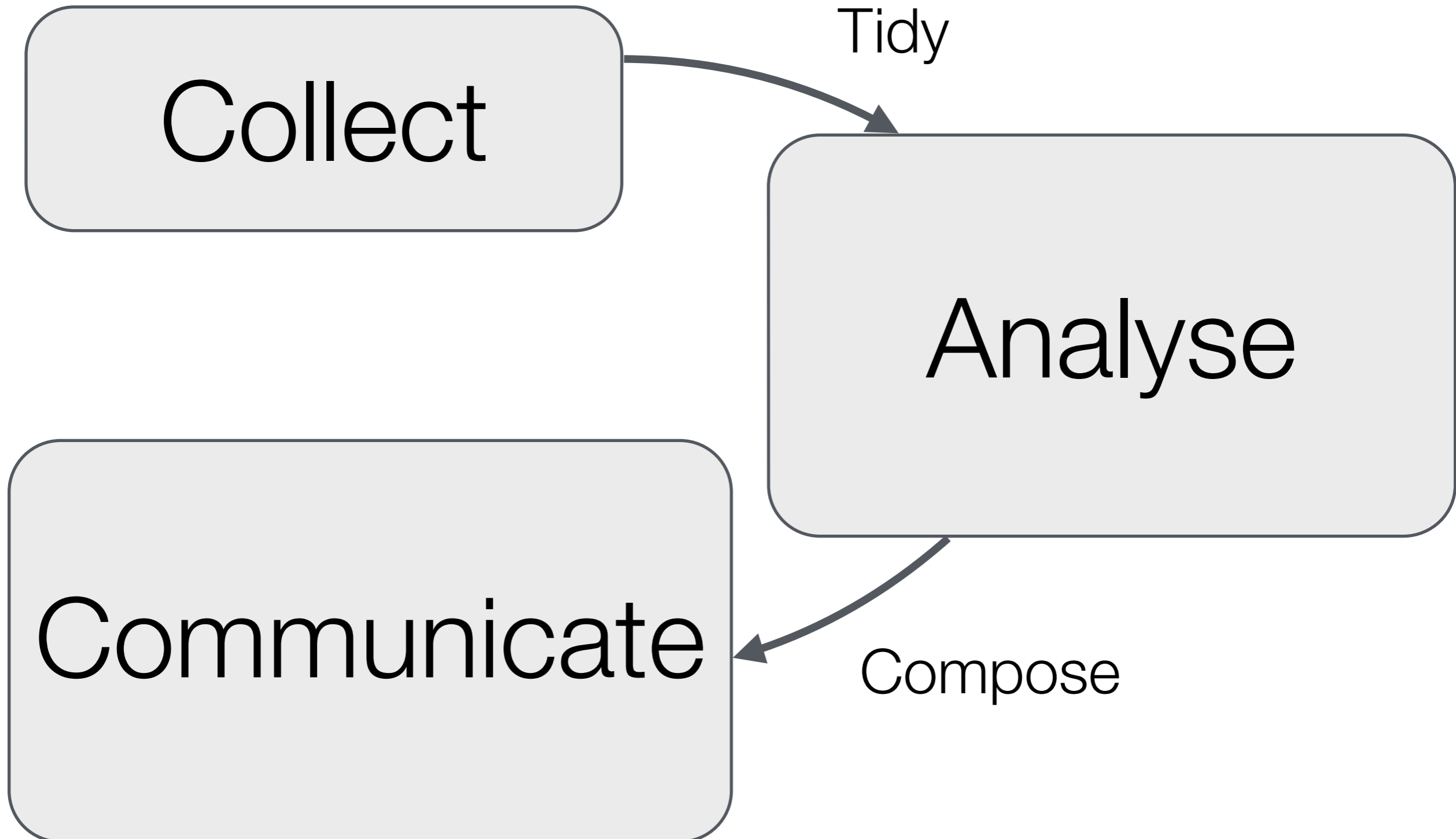
+

canvas

instructure.oregonstate.edu

We'll use canvas for handing in and grading work, taking quizzes, asking questions and announcements.
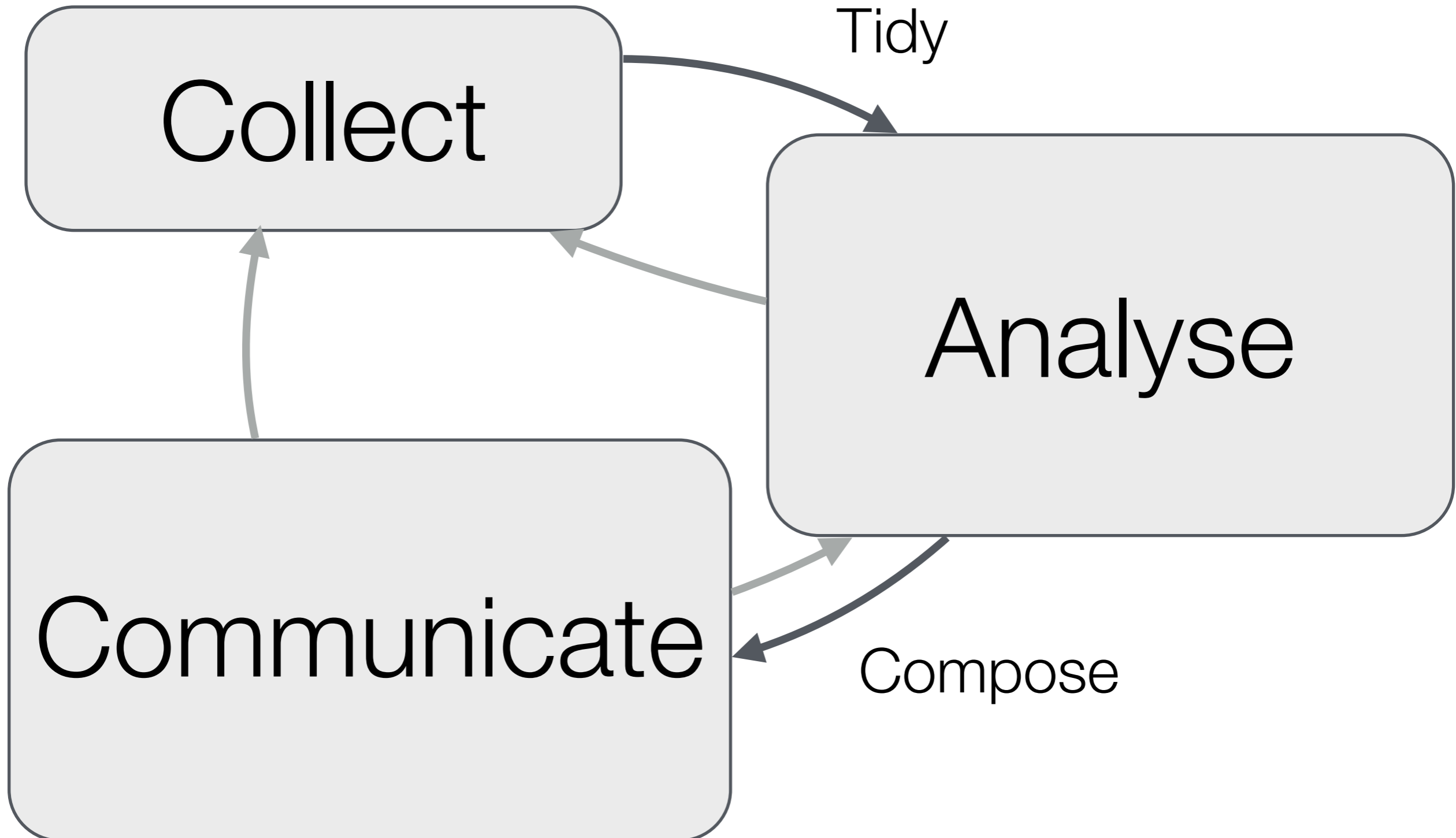
What do we learn in

ST411/511?

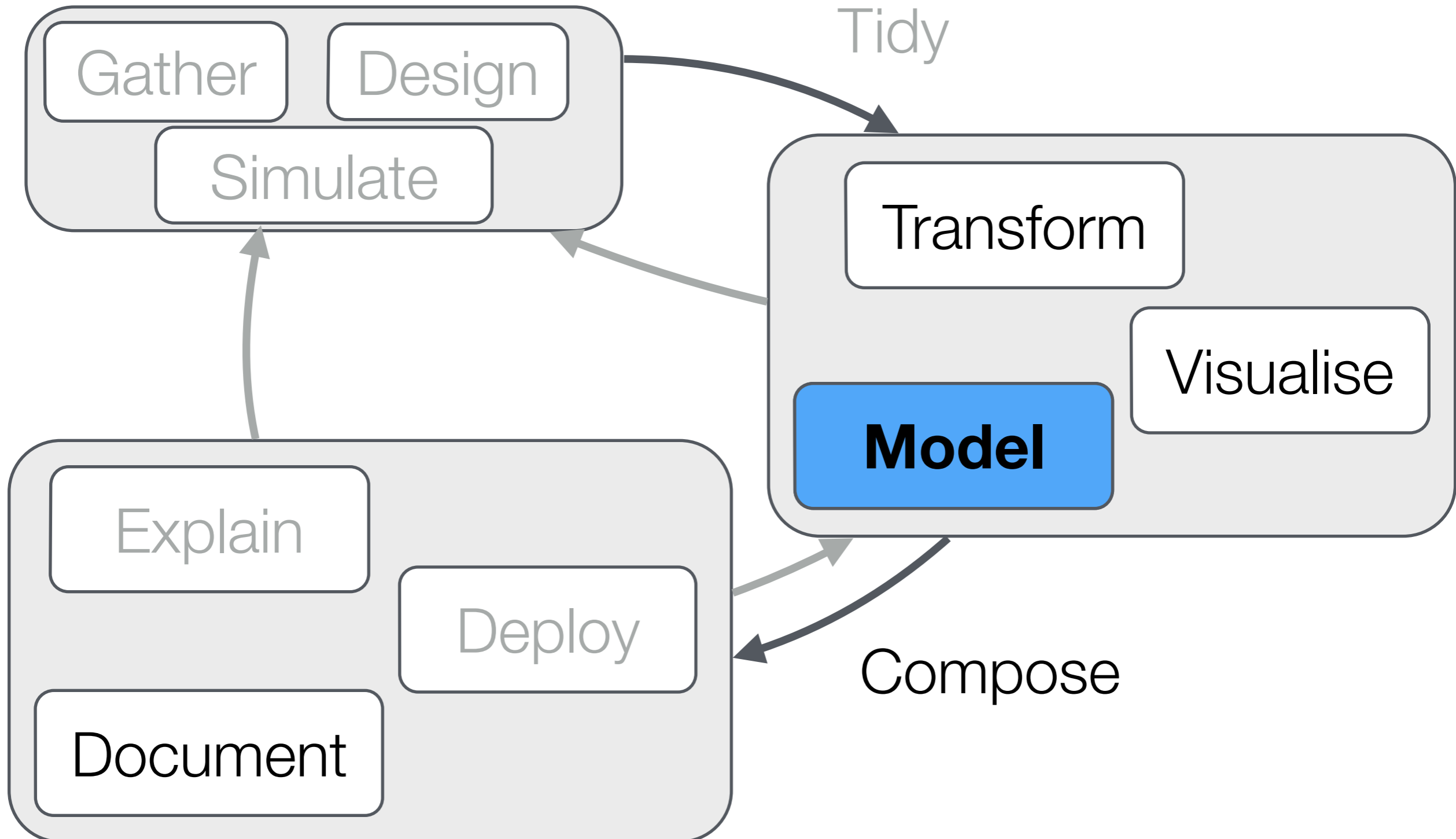# The **data** science process

starts with a question

The **data** science process

starts with a question

The **data** science process

starts with a question

Model   A statistical model describes how variability arises in our data.

Sometimes just knowing how the data was collected is enough to give you a model you can learn from.

Other times you have to assume a certain model, and convince yourself that at least the data doesn't grossly contradict that model.

The model usually has something unknown in it that you want to learn about.

We use the data we have, to learn about unknowns in the model.

Because there is variability in our data, we are always uncertain.

Statistics is all about quantifying that uncertainty.

Methods we will cover

Two group comparisons

Review t-tools, assumptions in depth, non-parametric tests

Multiple group comparisons

ANOVA, linear combination of means, multiple comparisons

Simple Linear regression

Chapters 1 - 8 in the textbook

More importantly,

Concepts we will cover

Hypothesis Testing

Estimation

Translating research questions to questions about statistical/probability models

The role and importance of assumptions

Writing non-technical (but statistically correct) summaries

Syllabus

stat511.cwick.co.nz

+

canvas

instructure.oregonstate.edu

We'll use canvas for handing in and grading work, taking quizzes, asking questions and announcements.

Prerequisites

ST 202 or ST 351
(this means an Introductory Statistics course)

You should be familiar with:
p-values
confidence intervals
the normal distribution
sampling distributions
simple random sampling
null hypotheses
doing a one and two sample t-test

You have a reading assignment for Monday,
it should feel familiar more on this later....

Textbook

The Statistical Sleuth by F. Ramsey and D. Shafer, 3rd Ed.

The 2nd Edition is also fine.

If you are buying used, you do not need the accompanying CD.

Readings will be assigned from the textbook.

Also available on reserve in the Library.

Assessment

Homework	20%
Data Analyses	30%
Quizzes	20%
Participation	10%
Final Exam	20%

Each week, either a homework or a data analysis will be due on Friday.

Homeworks are shorter, more directed and submitted as R notebooks.

Same for ST411 & ST511

Data analyses are longer, undirected and submitted as formal written reports.

Different for ST411 & ST511

Both are individual, you may discuss the assignments but your submission must be entirely your **own** work.

Homework 0

Your first R notebook (the format for homeworks)

Introduce yourself to me and your TA.

You will get started in lab.

Due on canvas before Friday Oct 2nd @11.59pm

Graded Complete/Incomplete

Quizzes will be roughly every two weeks in Blackboard.

They will be open for a limited time and timed

They will be multiple choice and short answer.

Quizzes are individual, you cannot discuss the content with anyone until the quiz has closed and been graded.

You **may** use your book and notes, but you shouldn't need to.

Approach them wisely:

- Study before taking the quiz
- If you refer to your notes for a question, study the topic some more after the quiz.
- Check the questions you get wrong, and study some more!
- Talk to me or the TA if you don't understand the answers

The **Final** exam will cover all material. Thursday Dec 10th at 6pm.

It will be a mixture of **quiz** type questions, and interpretation of analyses.

Participation grades will be based on study groups we form in week 4. More info later...

Academic Integrity

If I say something is "individual" I mean it.

Identical (or almost identical) homeworks or data analyses will result immediately in an **Academic Dishonesty Report**

Read the syllabus for other examples.

It's disrespectful to me, to your fellow students, and it devalues the worth of an Oregon State University degree.

If something outside school is impacting your ability to complete your work, talk to me.

Labs

Self-paced run by the TAs

Learn and practice R

Attendance is not required, and you don't have to submit anything.

Do NOT expect your TA to help you cover lab material if you chose not to attend lab!

A great place

- form study groups

- to start homework

- ask questions

Start next week!

If you aren't registered

Show up to a lab anyway.

Take a laptop with RStudio and R on it.

Or, get friendly with someone in lab and share a computer.

I can't guarantee you a spot, but usually everyone gets in by the end of the first week.

Let me know Mon in week two if haven't got a spot and you still want one.

Getting help

If you have a question about **class material, homework, R code**

- **Ask a question on the canvas discussion board!** (This is highly recommended over emailing, you reach me and all 4 TAs, and others benefit from our answers)

- Ask a question in lab.

- Go to your TAs office hours. TBA

- Come to my office hours / help sessions. **TBA**

If you have a question about your **own research analysis**

http://www.stat.oregonstate.edu/content/student-consulting-services

If you need to discuss a **personal matter**, email me.

Questions on syllabus?

R & RStudio

Why R?

Language of statisticians
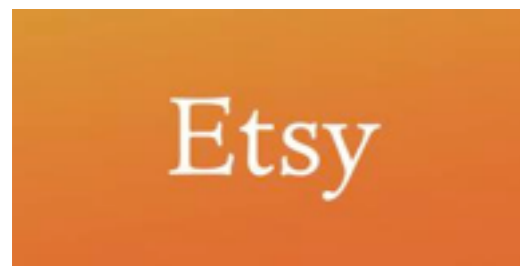
Becoming a common language for data analysis in all fields

Code is a form of communication

It's easy to extend

It's free

You'll be in good company

We'll be using R

- As a calculator

- To look up areas under curves

- To make plots

- To run statistical tests

This barely scratches the surface of what R can do

You will never be asked to write R code in a quiz or exam, but you could be asked to interpret R output.

What's RStudio?

R is the language.

RStudio is a place to write R code, run R code, and examine the results.

RStudio looks identical regardless of whether you are on Mac, Windows, Linux or using a web version.

You can use R & RStudio ...

In the basement labs in Milne
http://oregonstate.edu/is/mediaservices/scf

In Cordley 3003 (check schedule on door)

Or on your own computer

http://stat511.cwick.co.nz/installingRlocally.html

notebook-one.R

```r
# A histogram of a random sample of size 100
# from a Normal distribution with zero mean and
# variance one.


x <- rnorm(100)
qplot(x)


#' We can do basic formatting in the plain text
#' sections (the ones that start with #').
#' Extra hashes create headings
```

notebook-one.R

```r
# A histogram of a random sample of size 100
# f            stribution with zero mean and
# variance one.

x <-        00)
qplot(x)

#' We can do basic formatting in the plain text
#' s      the ones that start with #').
#' Extra hashes create headings
```

Code comment

R Code

Plain text

Time for a demo?

Next week

Review of basic statistical concepts.

Before Monday

Read: http://www.nature.com/nmeth/journal/v10/n9/full/nmeth.2613.html

Highlight statistical terminology. Look up anything you've forgotten and **bring** the article to class.

Install R and RStudio if you plan on using your own computer.

Completely free "Intro Stat" textbook:

https://www.openintro.org/stat/textbook.php?stat_book=os