Week 1: Getting Started

Alex Lishinski

August 19 2021

Agenda

1. Introductions to one another and the class
2. A brief introduction to data science and R
3. Discussion of the syllabus

Introductions!

- Alex Lishinski, Ph.D. (they/them)
- Contact:
  - alishins@utk.edu
- Postdoctoral researcher, CS Education, University of Tennessee, Knoxville
- Primary areas of interest:
  - Computer Science education
  - Quantitative research methods
  - Data science in education
- Former philosopher

Introductions!

1. Stuff about you
2. Why are you interested in data science?
3. What tools/experiences do you have for doing data science?
4. What do you think you want to get out of this course?

(10 minutes)

Course FAQ

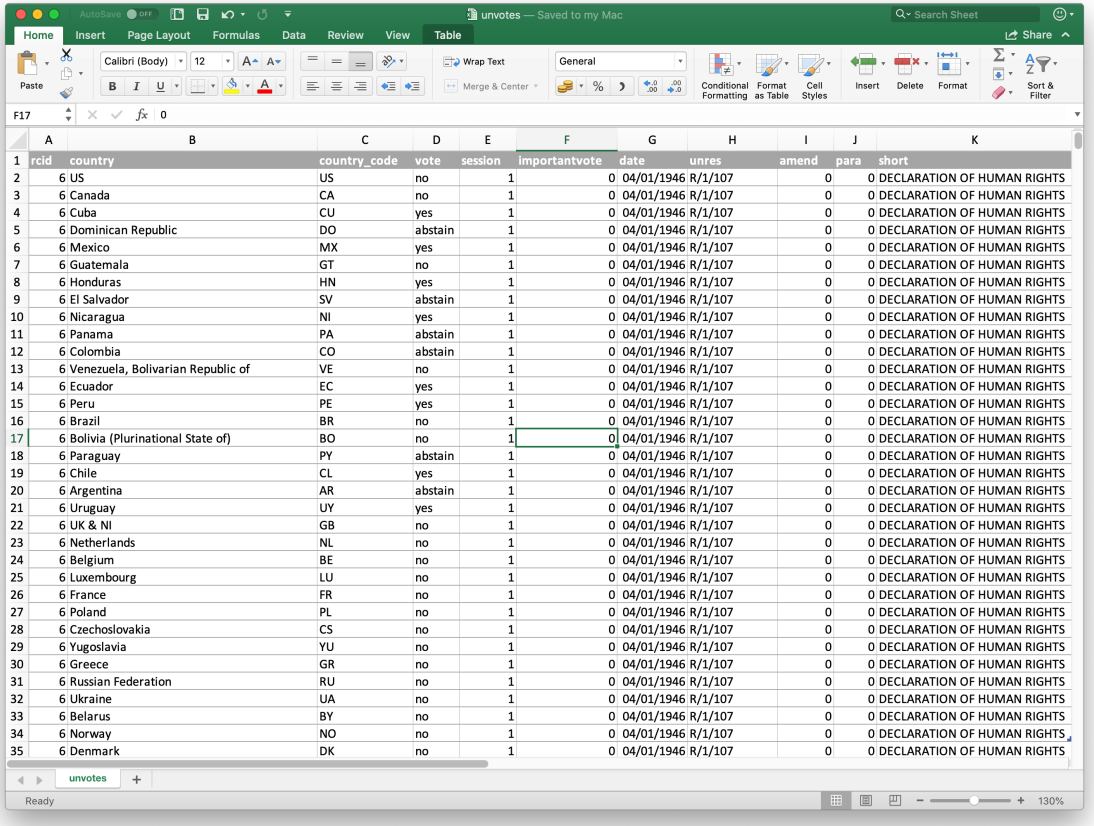
Q - What data science background does this course assume?
A - None.

Q - Is this an intro stat course?
A - While statistics $\neq$ data science, they are very closely related and have tremendous of overlap. Hence, this course is a great way to get started with statistics. However this course is *not* your typical high school statistics course.

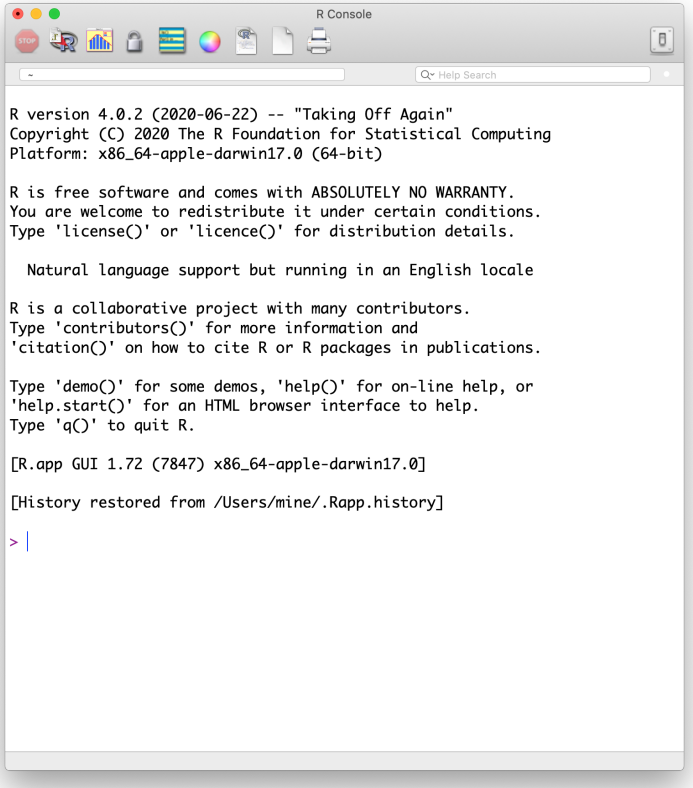
Q - Will we be doing computing?
A - Yes.

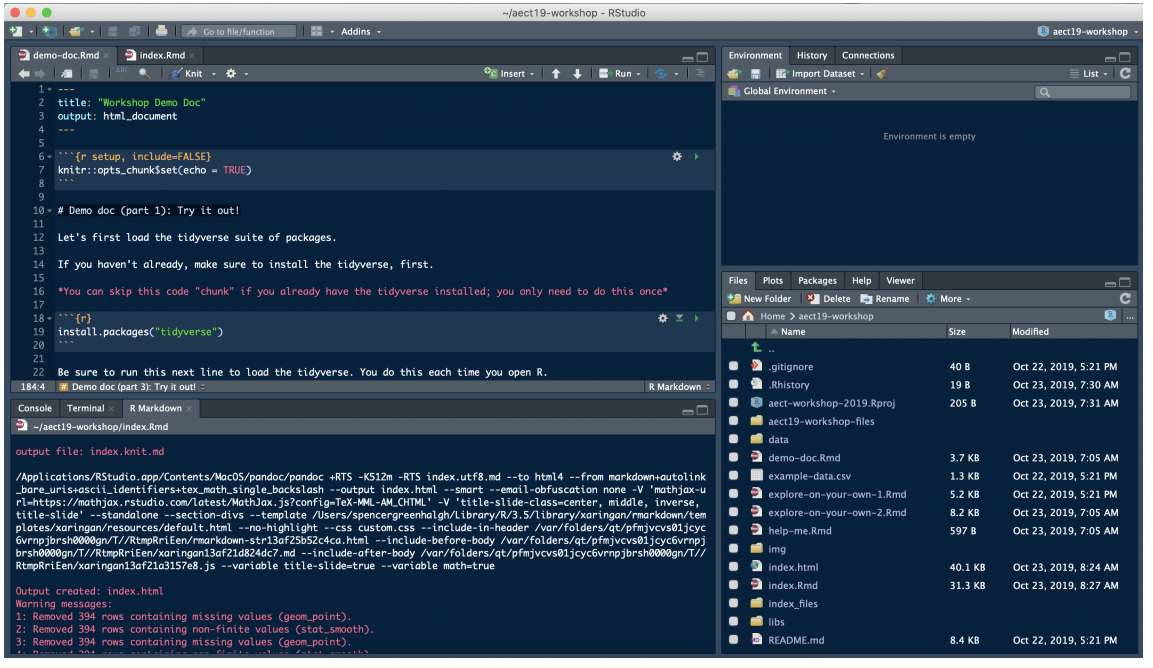
Q - Is this an intro CS course?
A - No, but many themes are shared.

Q - What computing language will we learn?
A - R.

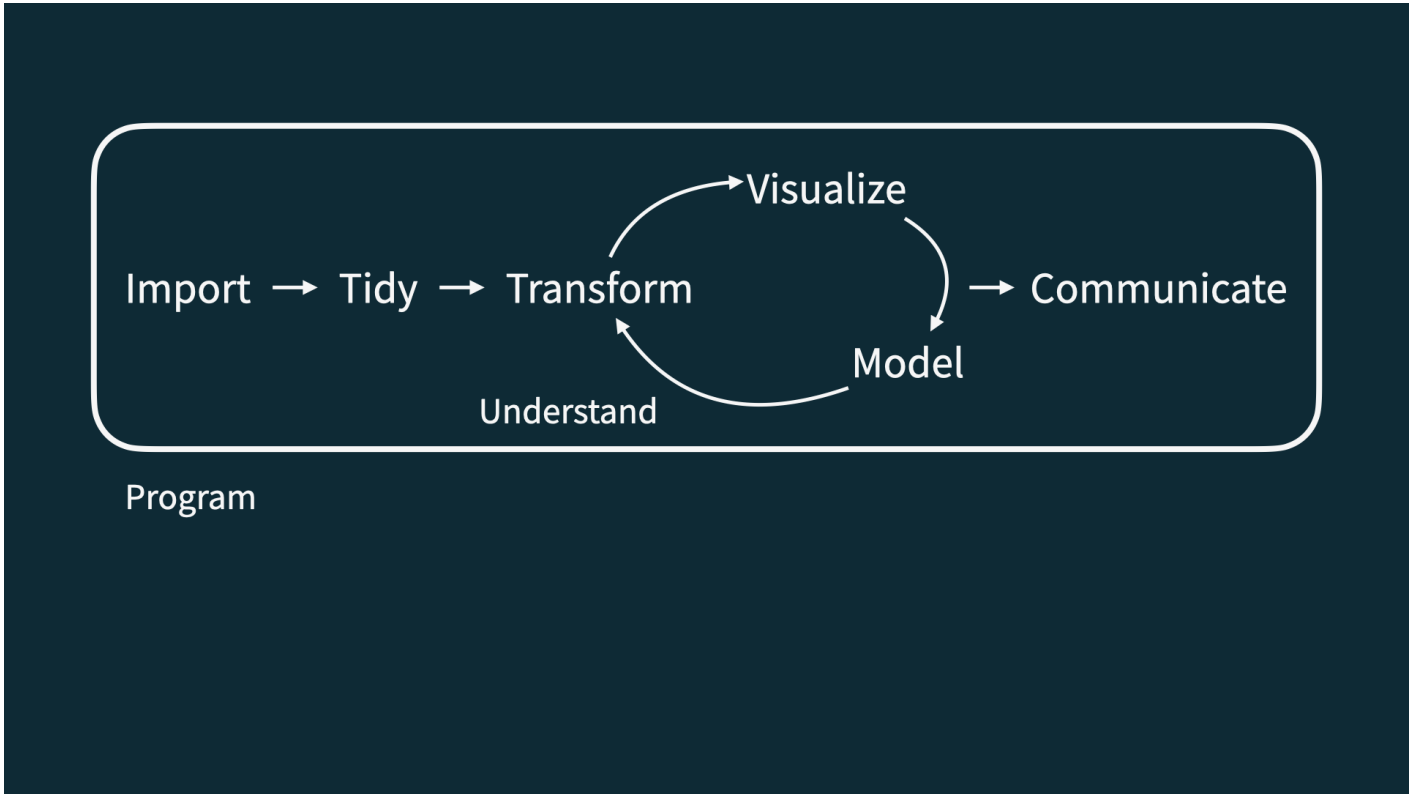


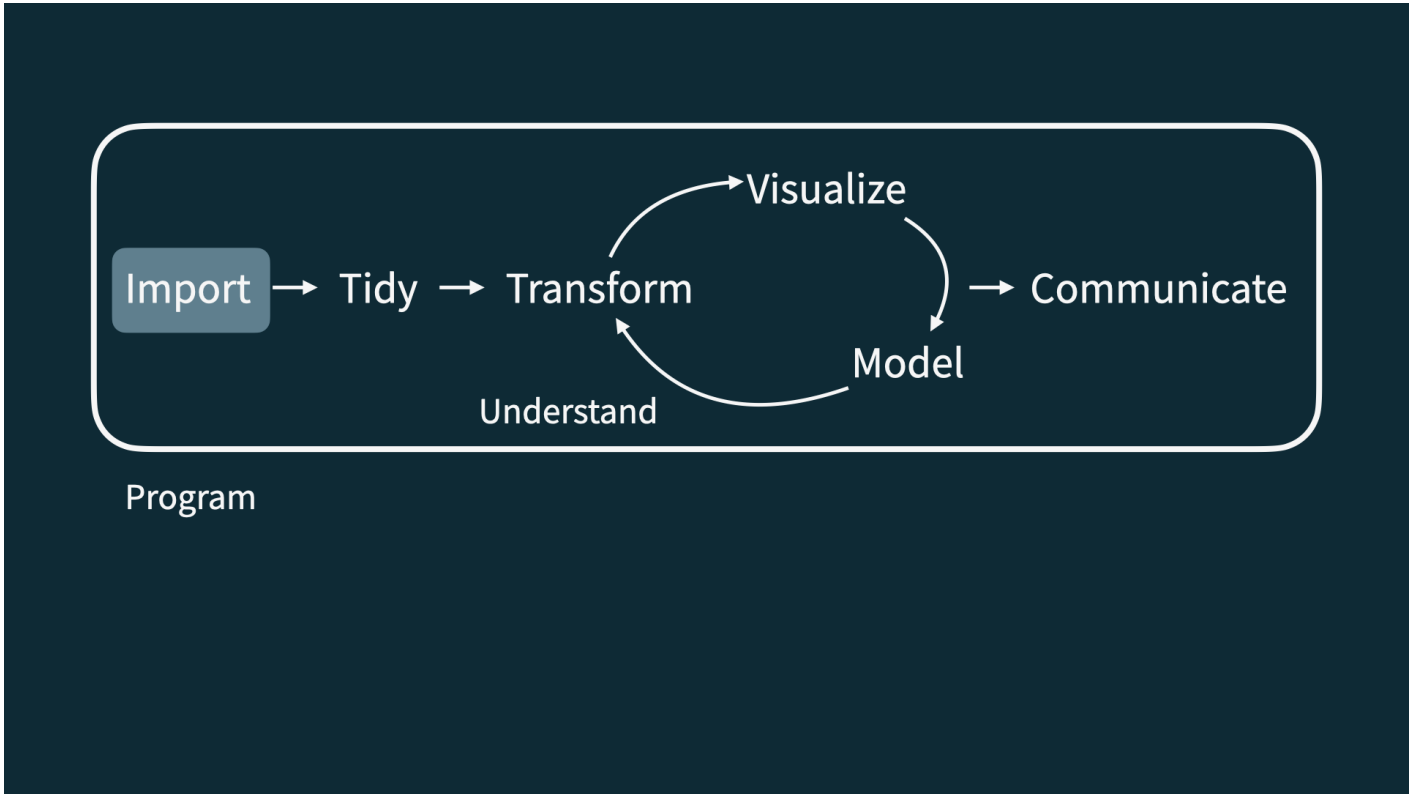
	A	B	C	D	E	F	G	H	I	J	K
1	rcid	country	country_code	vote	session	importantvote	date	unres	amend	para	short
2	6	US	US	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
3	6	Canada	CA	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
4	6	Cuba	CU	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
5	6	Dominican Republic	DO	abstain	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
6	6	Mexico	MX	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
7	6	Guatemala	GT	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
8	6	Honduras	HN	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
9	6	El Salvador	SV	abstain	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
10	6	Nicaragua	NI	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
11	6	Panama	PA	abstain	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
12	6	Colombia	CO	abstain	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
13	6	Venezuela, Bolivarian Republic of	VE	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
14	6	Ecuador	EC	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
15	6	Peru	PE	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
16	6	Brazil	BR	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
17	6	Bolivia (Plurinational State of)	BO	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
18	6	Paraguay	PY	abstain	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
19	6	Chile	CL	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
20	6	Argentina	AR	abstain	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
21	6	Uruguay	UY	yes	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
22	6	UK & NI	GB	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
23	6	Netherlands	NL	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
24	6	Belgium	BE	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
25	6	Luxembourg	LU	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
26	6	France	FR	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
27	6	Poland	PL	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
28	6	Czechoslovakia	CS	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
29	6	Yugoslavia	YU	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
30	6	Greece	GR	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
31	6	Russian Federation	RU	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
32	6	Ukraine	UA	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
33	6	Belarus	BY	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
34	6	Norway	NO	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS
35	6	Denmark	DK	no	1	0	04/01/1946	R/1/107	0	0	DECLARATION OF HUMAN RIGHTS

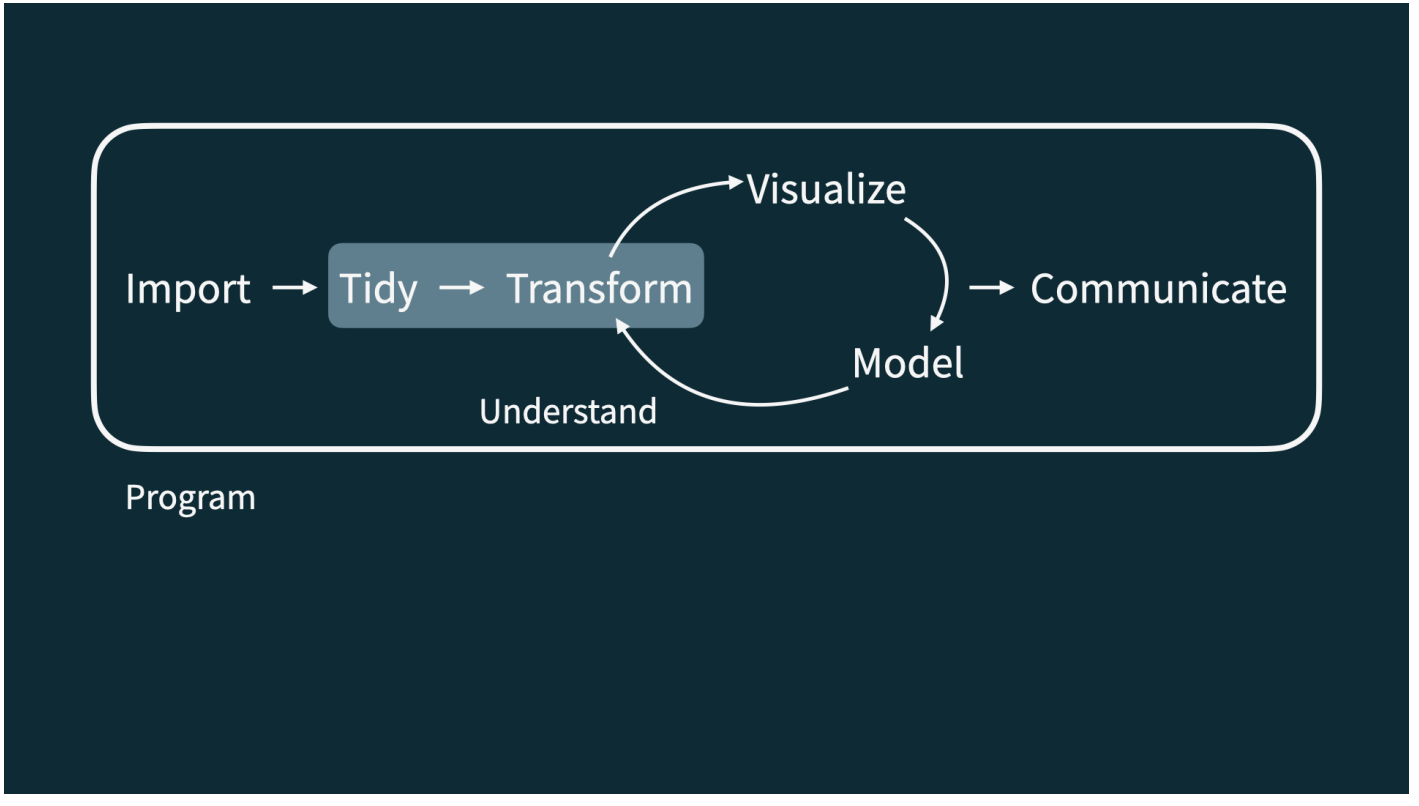

R Console
~
Help Search
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type '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.
[R.app GUI 1.72 (7847) x86_64-apple-darwin17.0]
[History restored from /Users/mine/.Rapp.history]
>

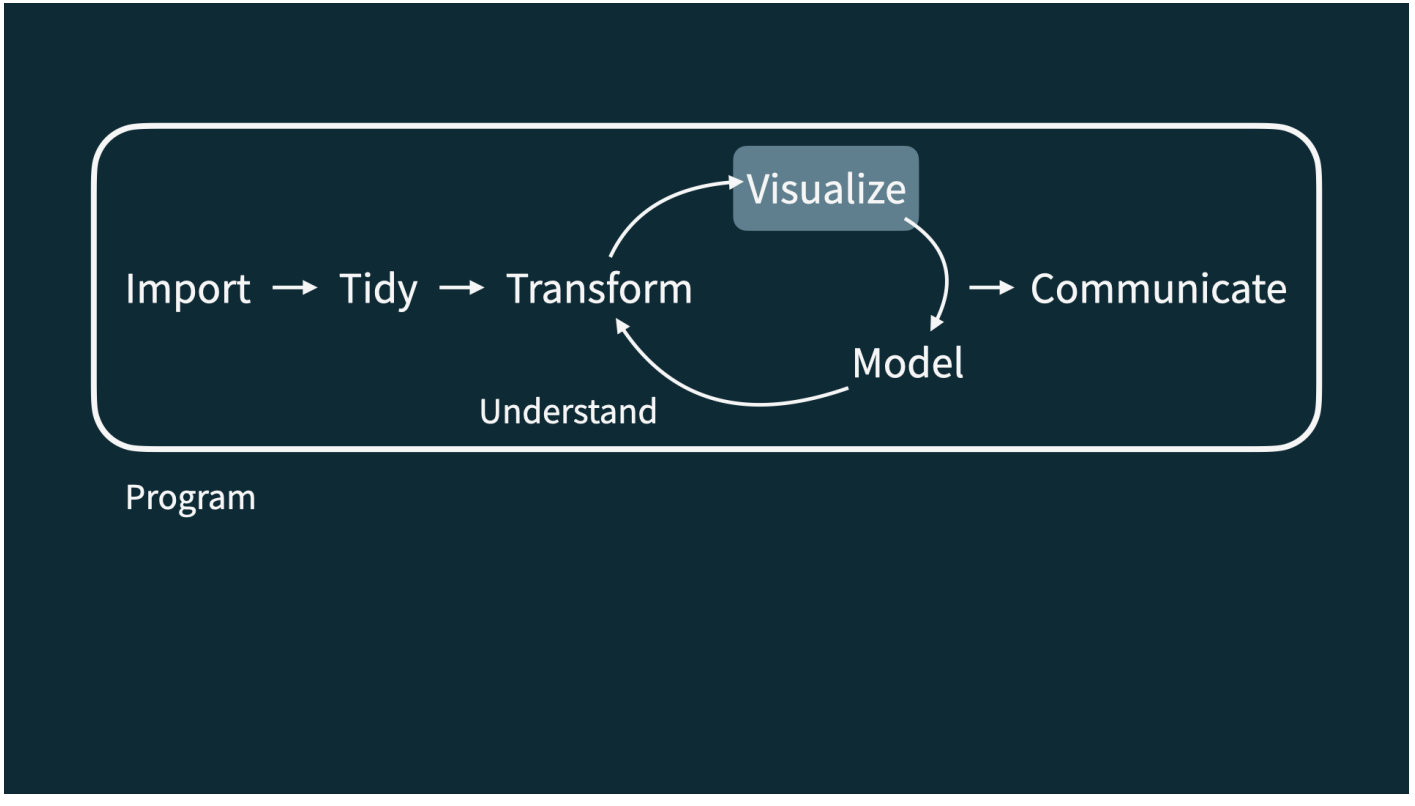

~/aect19-workshop - RStudio
demo-doc.Rmd
index.Rmd
Knit
Insert
Run
---
title: "Workshop Demo Doc"
output: html_document

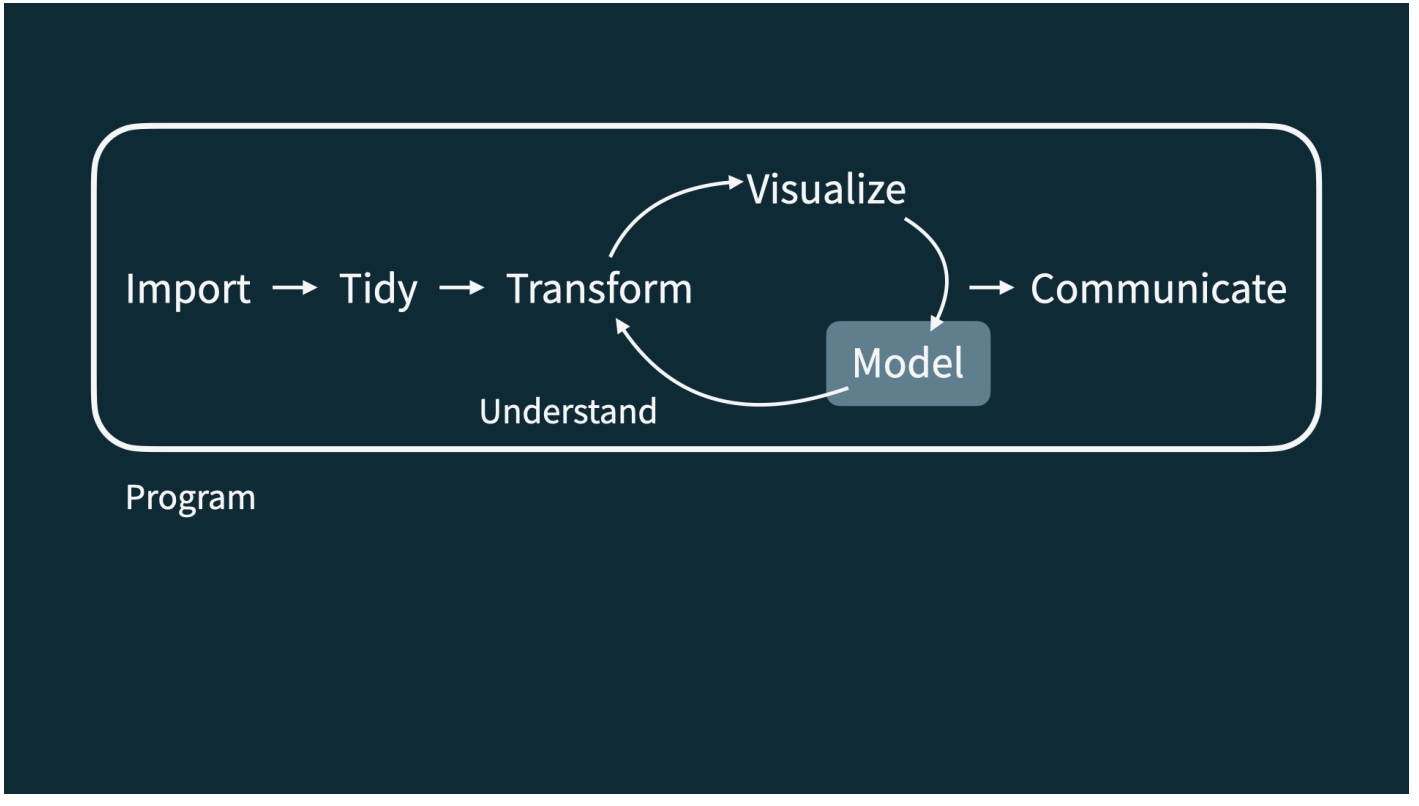
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
Demo doc (part 1): Try it out!
Let's first load the tidyverse suite of packages.
If you haven't already, make sure to install the tidyverse, first.
You can skip this code "chunk" if you already have the tidyverse installed; you only need to do this once
```{r}
install.packages("tidyverse")
```
Be sure to run this next line to load the tidyverse. You do this each time you open R.
184:4
Demo doc (part 3): Try it out!
R Markdown
Console
Terminal
R Markdown
~/aect19-workshop/index.Rmd
output file: index.knit.md
/Applications/RStudio.app/Contents/MacOS/pandoc/pandoc +RTS -K512m -RTS index.utf8.md --to html4 --from markdown+autolink_bare_uris+ascii_identifiers+tex_math_single_backslash --output index.html --smart --email-obfuscation none -V 'mathjax-url=https://mathjax.rstudio.com/latest/MathJax.js?config=TeX-MML-AM_CHTML' -V 'title-slide-class=center, middle, inverse, title-slide' --standalone --section-divs --template /Users/spencergreenhalgh/Library/R/3.5/library/xaringan/rmarkdown/templates/xaringan/resources/default.html --no-highlight --css custom.css --include-in-header /var/folders/qt/pfmjvcvs01jcyc6vrnpjbrsh0000gn/T//RtmpRriEen/rmarkdown-str13af25b52c4ca.html --include-before-body /var/folders/qt/pfmjvcvs01jcyc6vrnpjbrsh0000gn/T//RtmpRriEen/xaringan13af21d824dc7.md --include-after-body /var/folders/qt/pfmjvcvs01jcyc6vrnpjbrsh0000gn/T//RtmpRriEen/xaringan13af21a3157e8.js --variable title-slide=true --variable math=true
Output created: index.html
Warning messages:
1: Removed 394 rows containing missing values (geom_point).
2: Removed 394 rows containing non-finite values (stat_smooth).
3: Removed 394 rows containing missing values (geom_point).
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home > aect19-workshop
Name
Size
Modified
.gitignore 40 B Oct 22, 2019, 5:21 PM
.Rhistory 19 B Oct 23, 2019, 7:30 AM
aect-workshop-2019.Rproj 205 B Oct 23, 2019, 7:31 AM
aect19-workshop-files
data
demo-doc.Rmd 3.7 KB Oct 23, 2019, 7:05 AM
example-data.csv 1.3 KB Oct 22, 2019, 5:21 PM
explore-on-your-own-1.Rmd 5.2 KB Oct 22, 2019, 5:21 PM
explore-on-your-own-2.Rmd 8.2 KB Oct 23, 2019, 7:05 AM
help-me.Rmd 597 B Oct 23, 2019, 7:05 AM
img
index.html 40.1 KB Oct 23, 2019, 8:24 AM
index.Rmd 31.3 KB Oct 23, 2019, 8:27 AM
index_files
libs
README.md 8.4 KB Oct 22, 2019, 5:21 PM

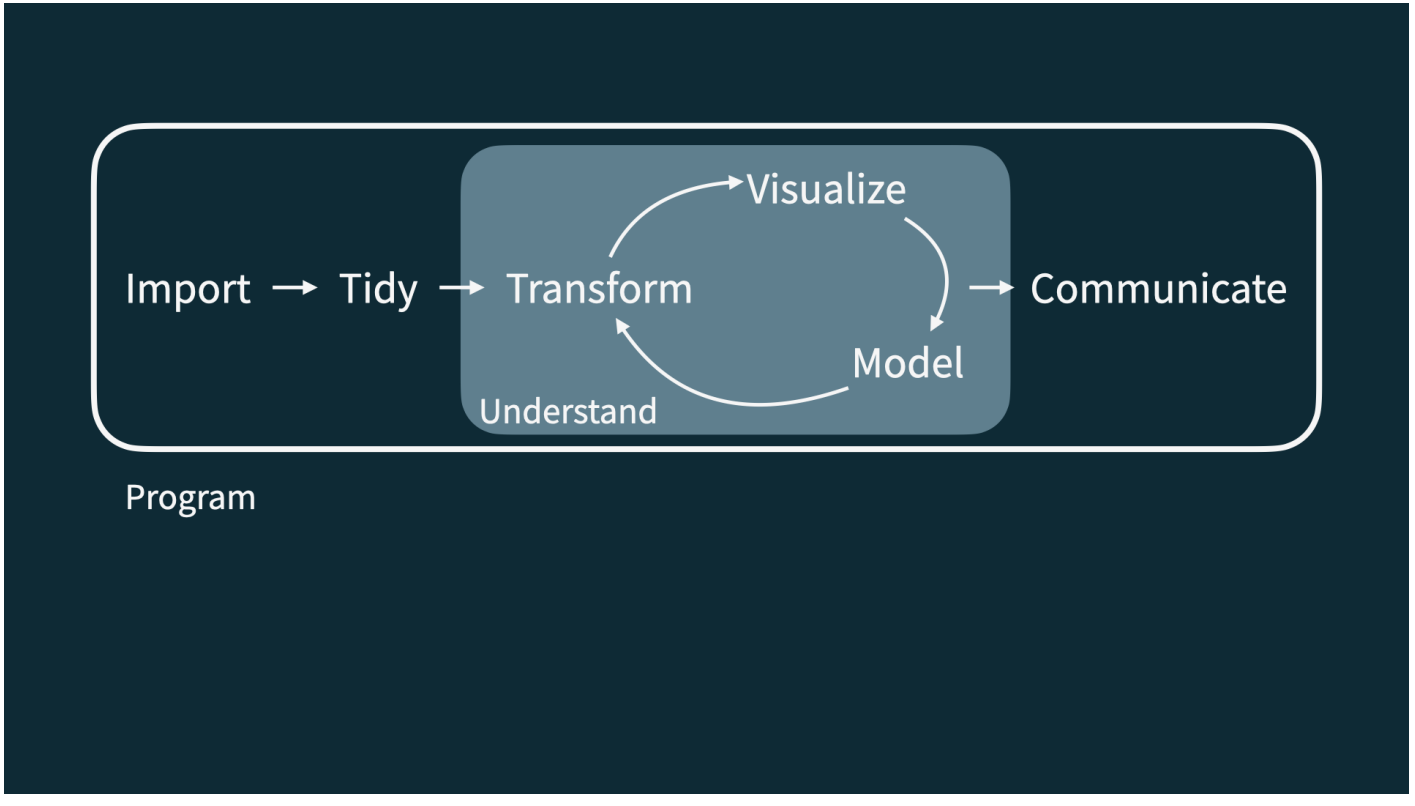

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

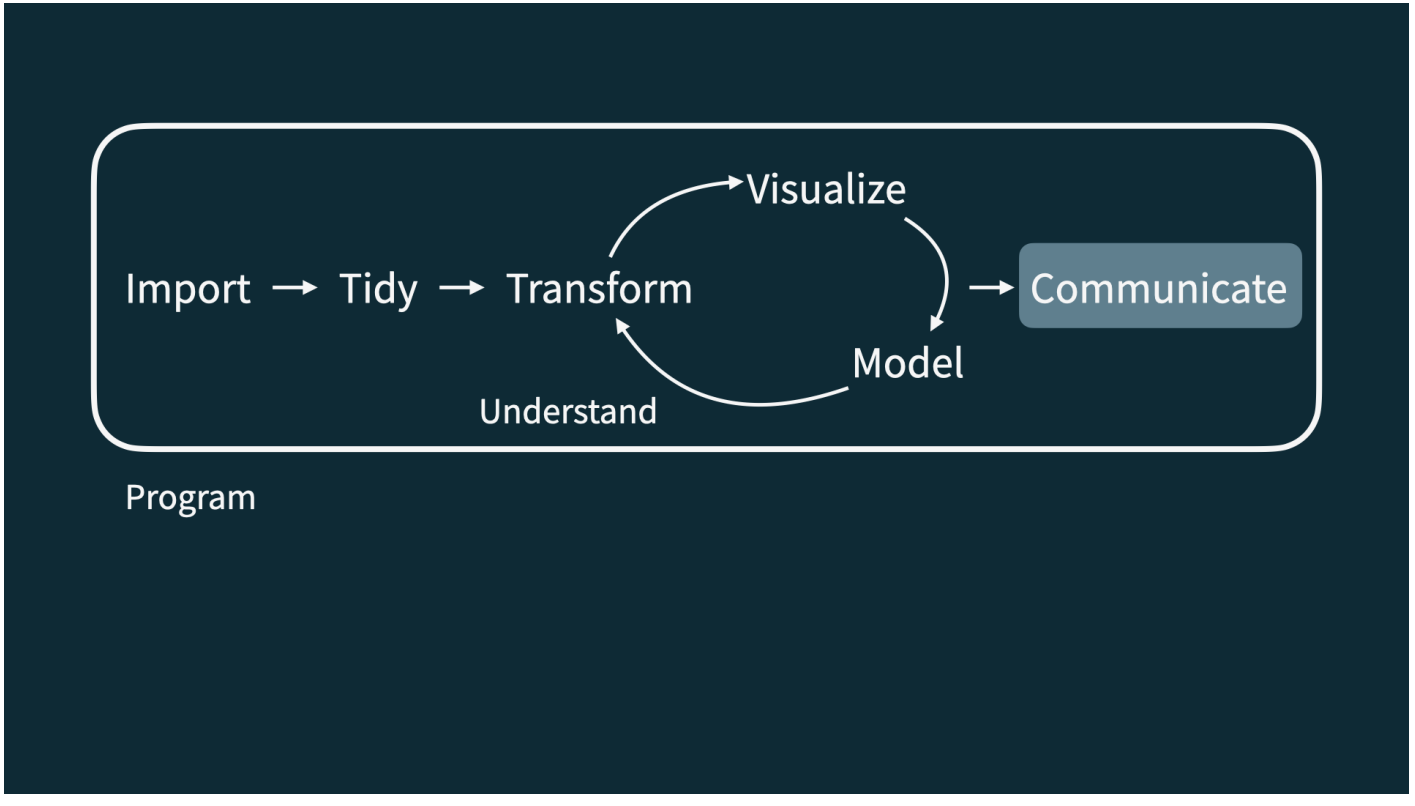

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

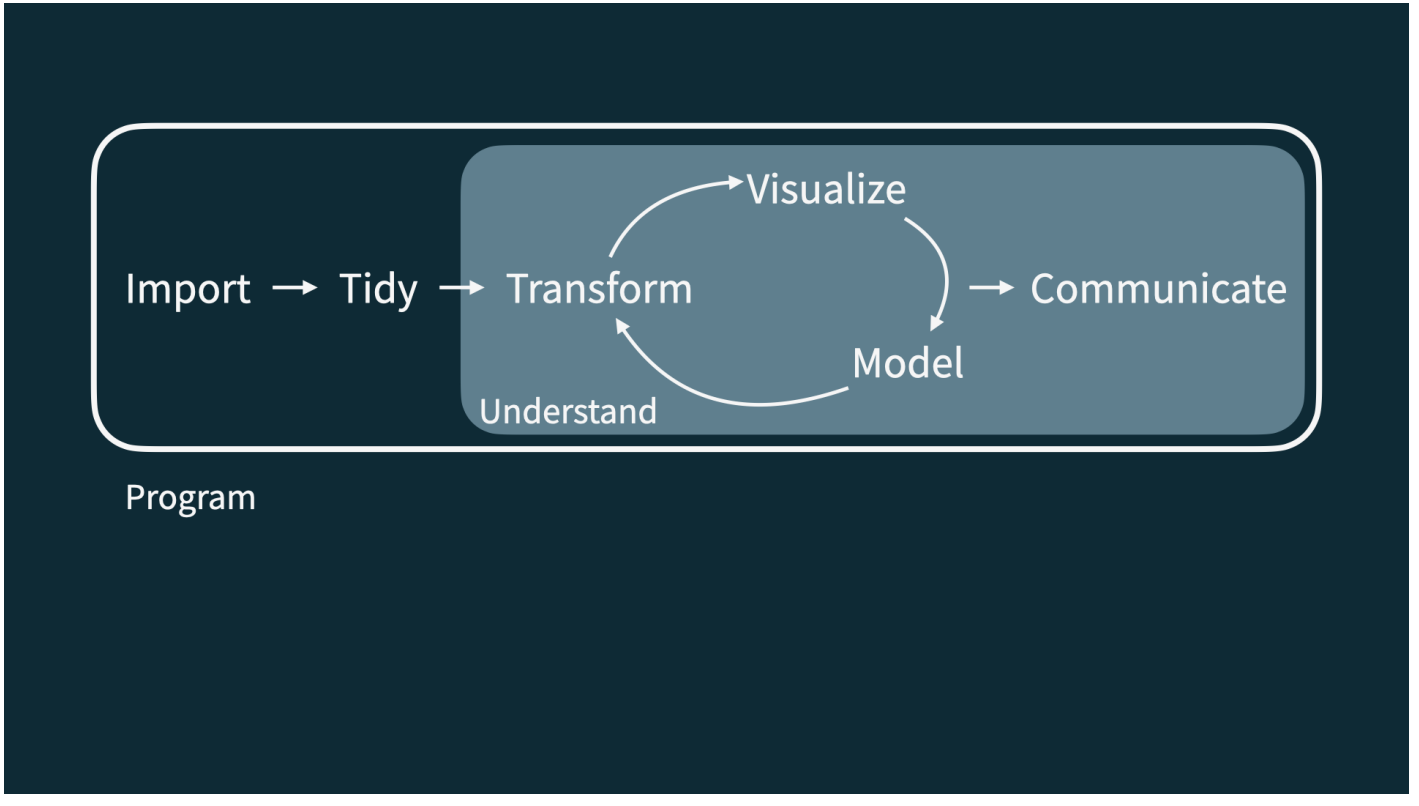

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

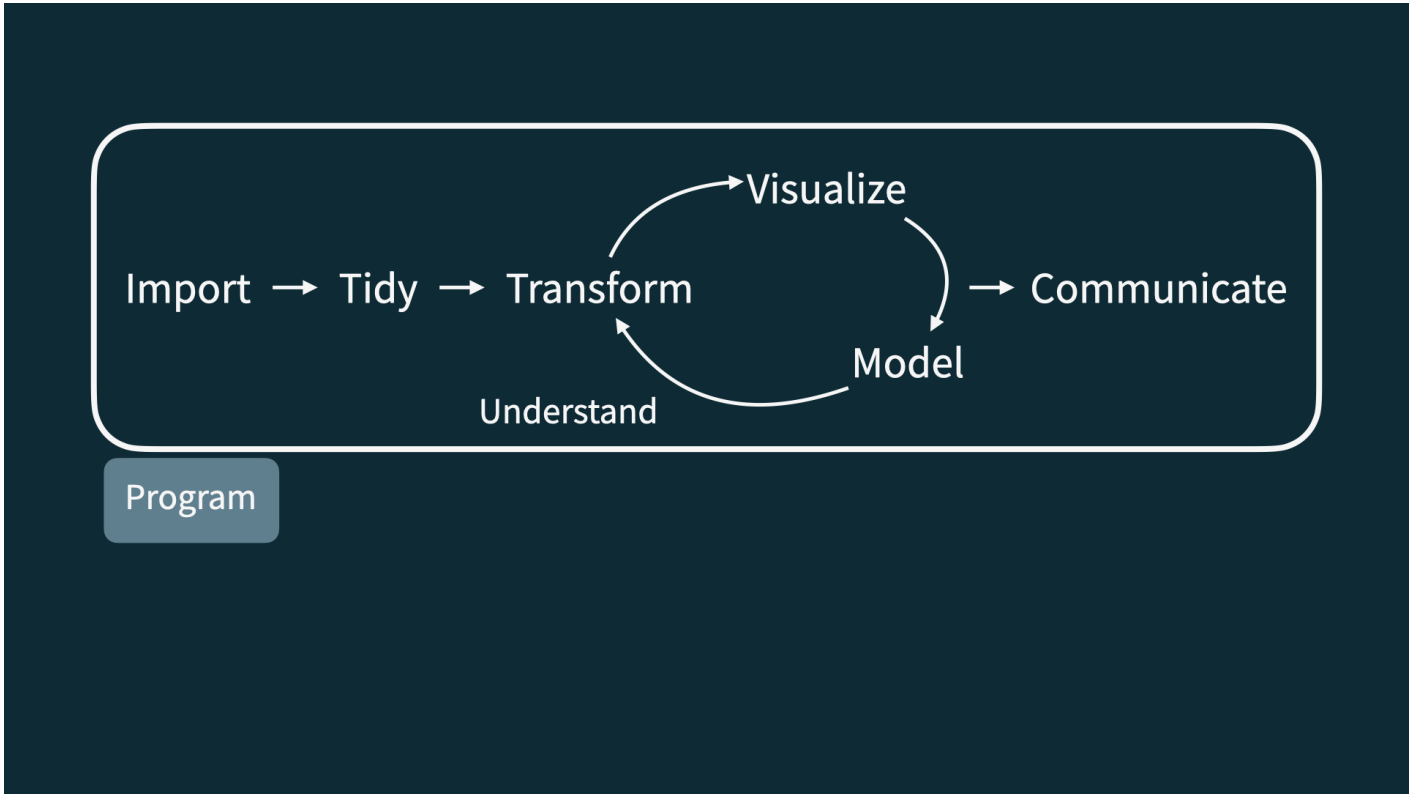

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program


Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program


Import
Tidy
Transform
Visualize
Model
Understand
Communicate
Program


Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program


Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program


Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

Over-arching design

The pedagogical framework ***plays out*** in the following specific ways:

1. You'll first hear about the concepts underlying what you can do
2. You'll see examples of working out these ideas in practice
3. You'll try these ideas out on your own

An argument for this class

- This presentation will make the argument that **you can use R and RStudio to solve problems that are important to you**
- In doing so, you can **participate in a cool community of data scientists in education**

Why learn R?

- It is capable of carrying out basic and complex statistical analyses
- It is able to work with data small ($n = 30$) and large ($n = 100{,}000+$) efficiently
- It is a programming language and so is quite flexible
- There is a great, inclusive community of users and developers (and teachers)
- It is increasingly used in education
- It can help you to carry out your educational analyses in open and trustworthy ways
- It is cross-platform, open-source, and freely-available

RMarkdown

- RMarkdown is a data analysis "notebook" that combines text with code and output
- It is a great file type to use when beginning to use R and to create reproducible analyses
- It is fun to use because you can generate different types of output (Word, PDF, and even web-based)

Let's look at a bit of code together

What do you think this code will do?

```
sci_mo_processed %>%
  filter(percentage_earned >= .60) %>%
  select(student_id, course_id, percentage_earned))
```

Let's look at a bit of code together

```
sci_mo_processed %>%
  filter(percentage_earned >= .60) %>%
  select(student_id, course_id, percentage_earned)
```

```
## # A tibble: 563 × 3
##    student_id course_id     percentage_earned
##         <dbl> <chr>                     <dbl>
##  1      43146 FrScA-S216-02             0.677
##  2      44638 OcnA-S116-01              0.757
##  3      47448 FrScA-S216-01             0.661
##  4      47979 OcnA-S216-01              0.677
##  5      48797 PhysA-S116-01             0.865
##  6      51943 FrScA-S216-03             0.855
##  7      52446 PhysA-S116-01             0.824
##  8      53447 FrScA-S116-01             0.676
##  9      53475 FrScA-S116-02             0.820
## 10      53475 FrScA-S216-01             0.808
## # … with 553 more rows
```

Let's look at a bit of code together

What do you think this code will do?

```
sci_mo_processed %>%
  filter(percentage_earned >= .60) %>%
  arrange(desc(percentage_earned)) %>%
  select(student_id, course_id, percentage_earned, TimeSpent)
```

Let's look at a bit of code together

```
sci_mo_processed %>%
  filter(percentage_earned >= .60) %>%
  select(student_id, course_id, percentage_earned, TimeSpent)
```

```
## # A tibble: 563 × 4
##    student_id course_id     percentage_earned TimeSpent
##         <dbl> <chr>                     <dbl>     <dbl>
##  1      43146 FrScA-S216-02             0.677   1555.
##  2      44638 OcnA-S116-01              0.757   1383.
##  3      47448 FrScA-S216-01             0.661    860.
##  4      47979 OcnA-S216-01              0.677   1599.
##  5      48797 PhysA-S116-01             0.865   1482.
##  6      51943 FrScA-S216-03             0.855      3.45
##  7      52446 PhysA-S116-01             0.824   1390.
##  8      53447 FrScA-S116-01             0.676   1479.
##  9      53475 FrScA-S116-02             0.820     NA
## 10      53475 FrScA-S216-01             0.808   1867.
## # … with 553 more rows
```

Getting Started

- Download R: https://www.r-project.org/
- Download Rstudio: https://www.rstudio.com/products/rstudio/download/
- Set up a working directory for your class projects

Syllabus

Assignments

- Homework
- Exams
- Final Project
- Others (data ethics, data science resources)

Organization of the class

- Tuesdays - Homeworks due
- Tuesday class - Slides and demo
- Thursday class - Review HW and work through hands on activity

Questions

What questions do you have?