



Welcome to the Tidyverse

An Introduction To Data Science

Teon Brooks

Carl Howe

Christina Koch

and

A special mystery instructor



Introduction

HELLO

my name is

Carl

HELLO

my name is

Christina

HELLO

my name is

Charlotte

HELLO

my name is

Teon

Your Turn

Introduce yourselves to your neighbors:

Who are you?

Where are you from?

What made you sign up for today's workshop?

Our Day

9:00-9:30 AM	Welcome and Introduction
9:30-10:45 AM	Getting Started
10:45-11:00 AM	Break
11:00-12:00 AM	Data Basics
12:00-1:00 PM	Lunch
1:00-2:45 PM	Data Visualization
2:45-3:00 PM	Break
3:00-4:50 PM	Tidy Data and Data Transformation
4:50-5:00 PM	Summary and Wrap Up

What is Data Science?

R

Why Study Data Science?

R

GLASSDOOR SAYS ITS A LUCRATIVE JOB

Data Scientist Salaries

4,354 Salaries Updated Feb 4, 2019

About This Data ?

Industries lock Company Sizes lock Years of Experience lock

i To continue using salary filters, please contribute. [Write a Review](#) or [Add a Salary](#)

Average Base Pay

\$117,345 /yr



Additional Cash Compensation ?

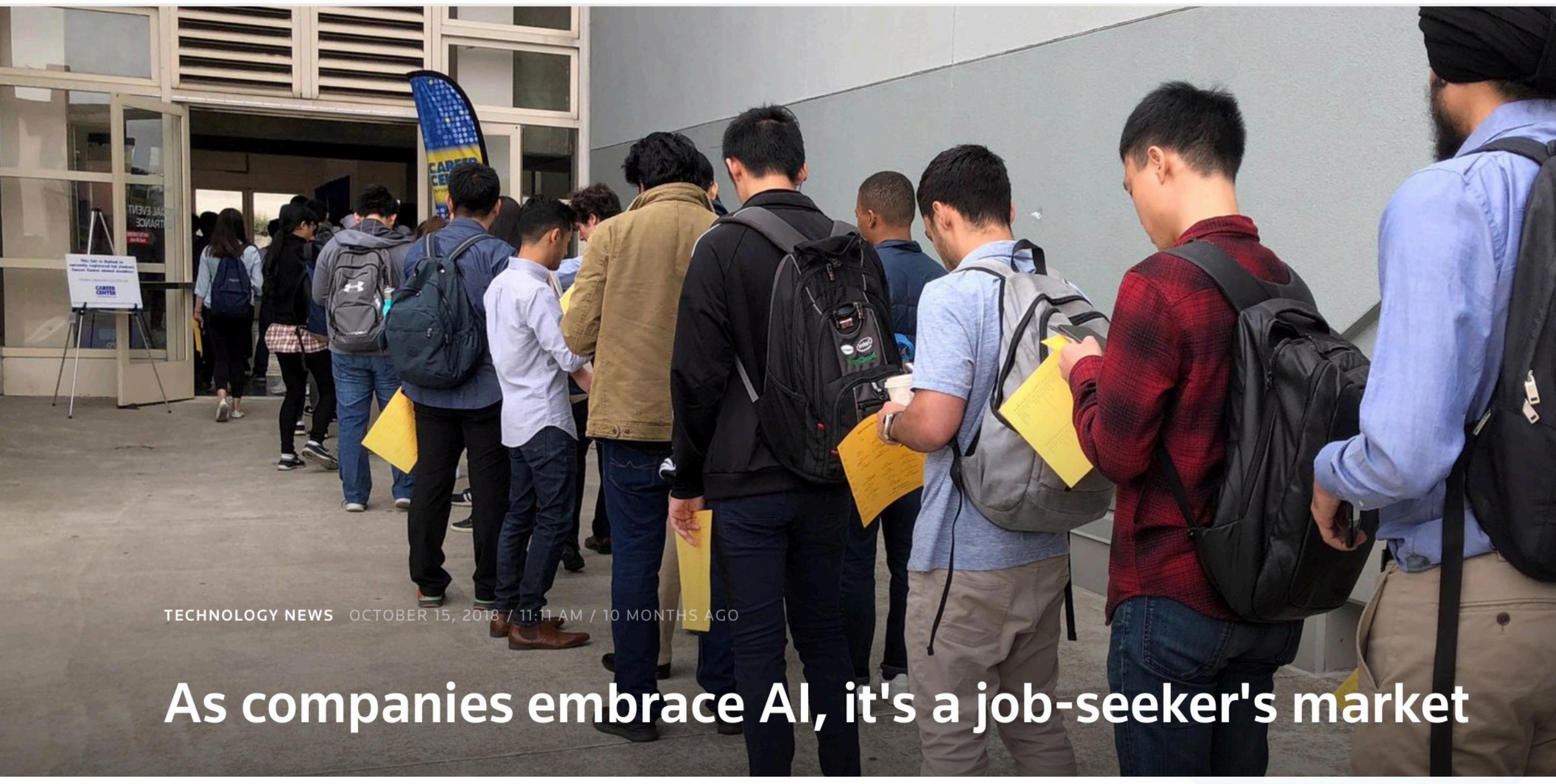
Average \$xx,xxx

Range \$xx,xxx

How much does a Data Scientist make?
The national average salary for a Data Scientist is \$117,345
in United States. Filter by location to see... [More](#)

Salaries for Related Job Titles

Data Analyst	\$67K
Data Scientist Intern	\$87K
Quantitative Analyst	\$116K
Senior Data Scientist	\$139K



TECHNOLOGY NEWS OCTOBER 15, 2018 / 11:11 AM / 10 MONTHS AGO

As companies embrace AI, it's a job-seeker's market

SingularityHub



Most highly valued company, at \$75B. It is also the most highly-funded unicorn on the list with \$3.1B in total equity funding



Fastest to unicorn, reaching a \$1B valuation with its 1st ever equity round of \$154M



Unicorns with least amount of disclosed funding: Momenta's funding, excluding an undisclosed Series C, is \$62M; Orcam raised \$77.4M

*includes startups offering AI-SaaS or using machine learning as a core differentiator for their main product offering; excludes AR/VR and hardware-focused robotics startups

- > Big Data with R Workshop
 - > Modern Geospatial Data Analysis with R Workshop
- > Deep Learning with Keras and TensorFlow in R Workflow
 - > Text Mining with Tidy Data Principles Workshop



*“The future's so bright, I
gotta wear shades!”*

Timbuk 3

Let's Do Some Data Science!



Complete the world's shortest mobile phone survey

<https://bit.ly/33hYRcj>

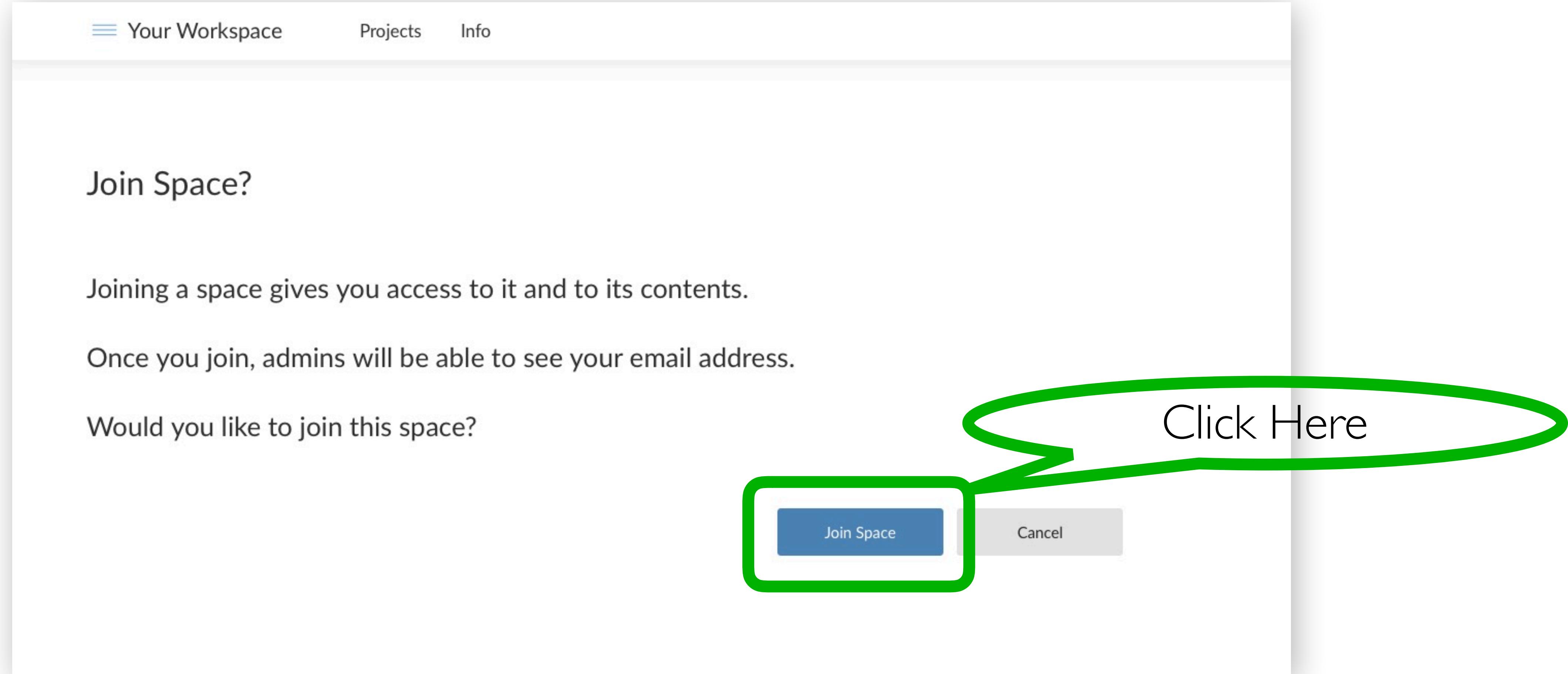
Now go to the following URL

<http://tidy-atl-cloud>

Log in using your rstudio.cloud
account you created earlier

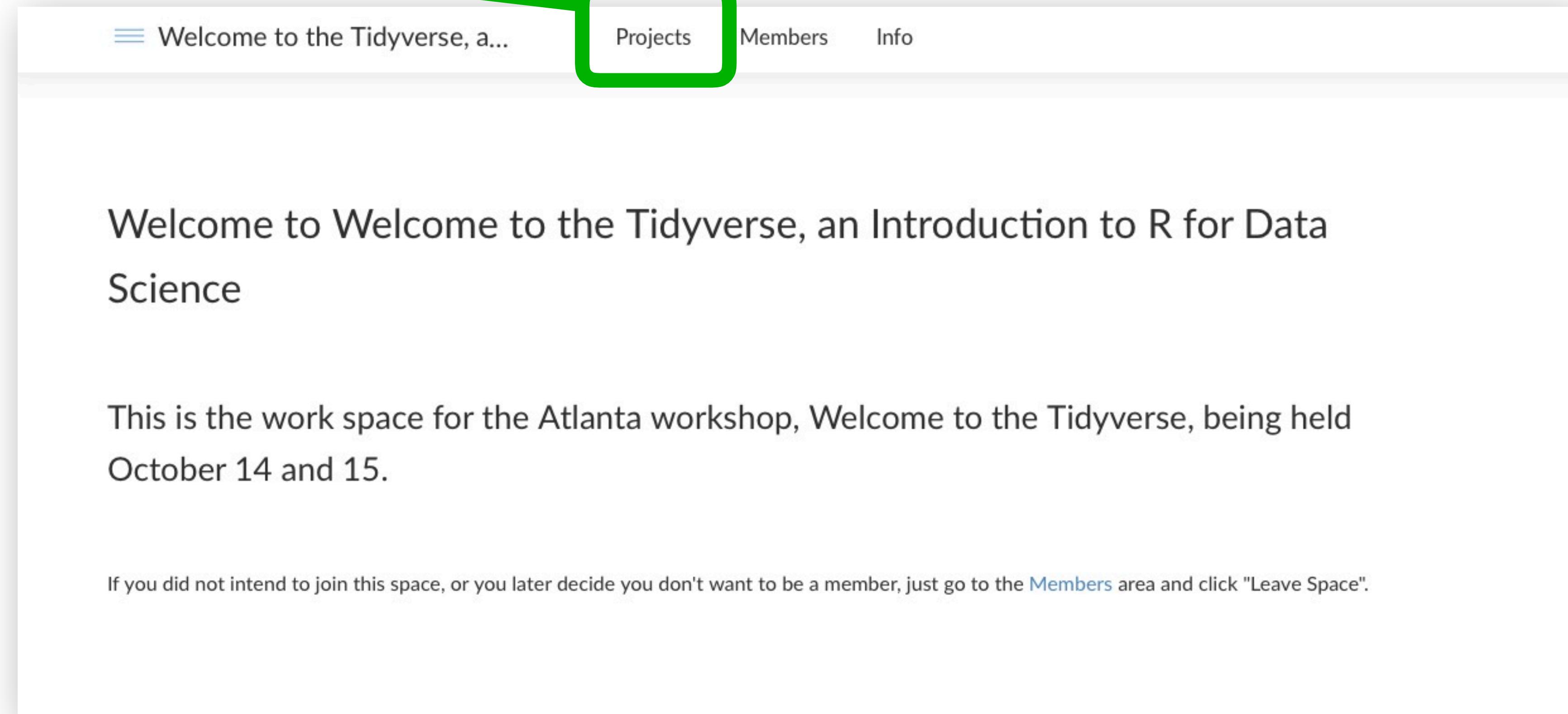
<http://tidy-atl-cloud>

You should see this screen:

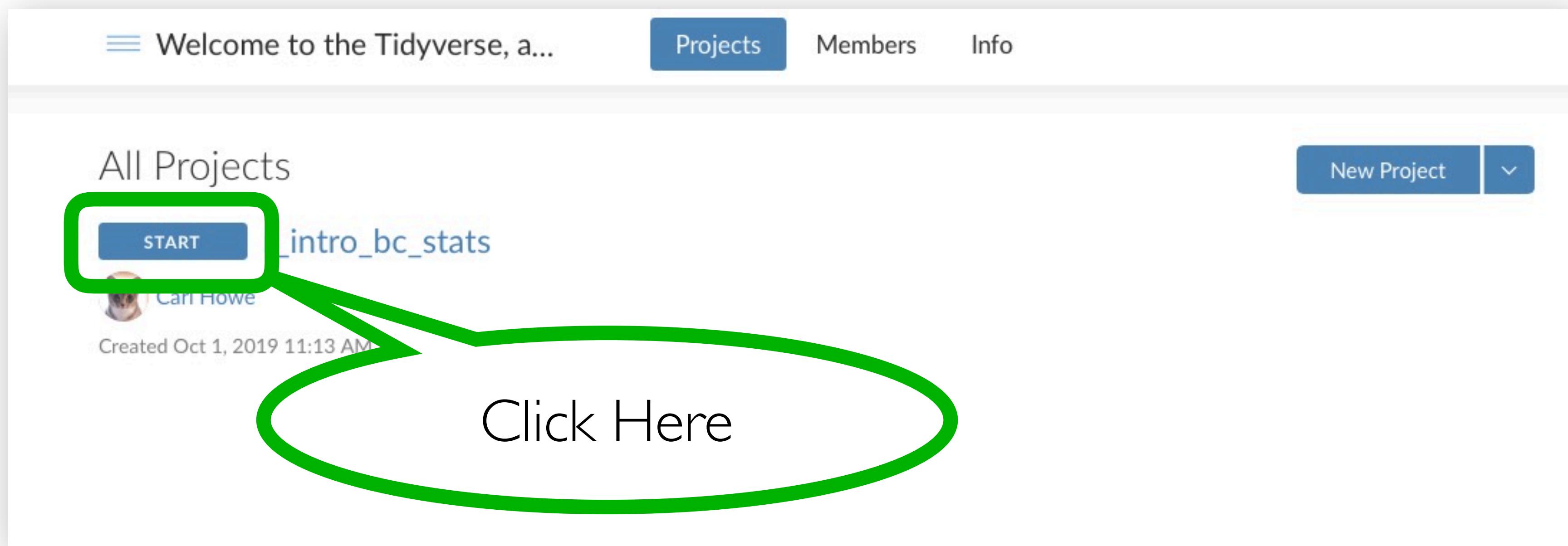


After you join you see this

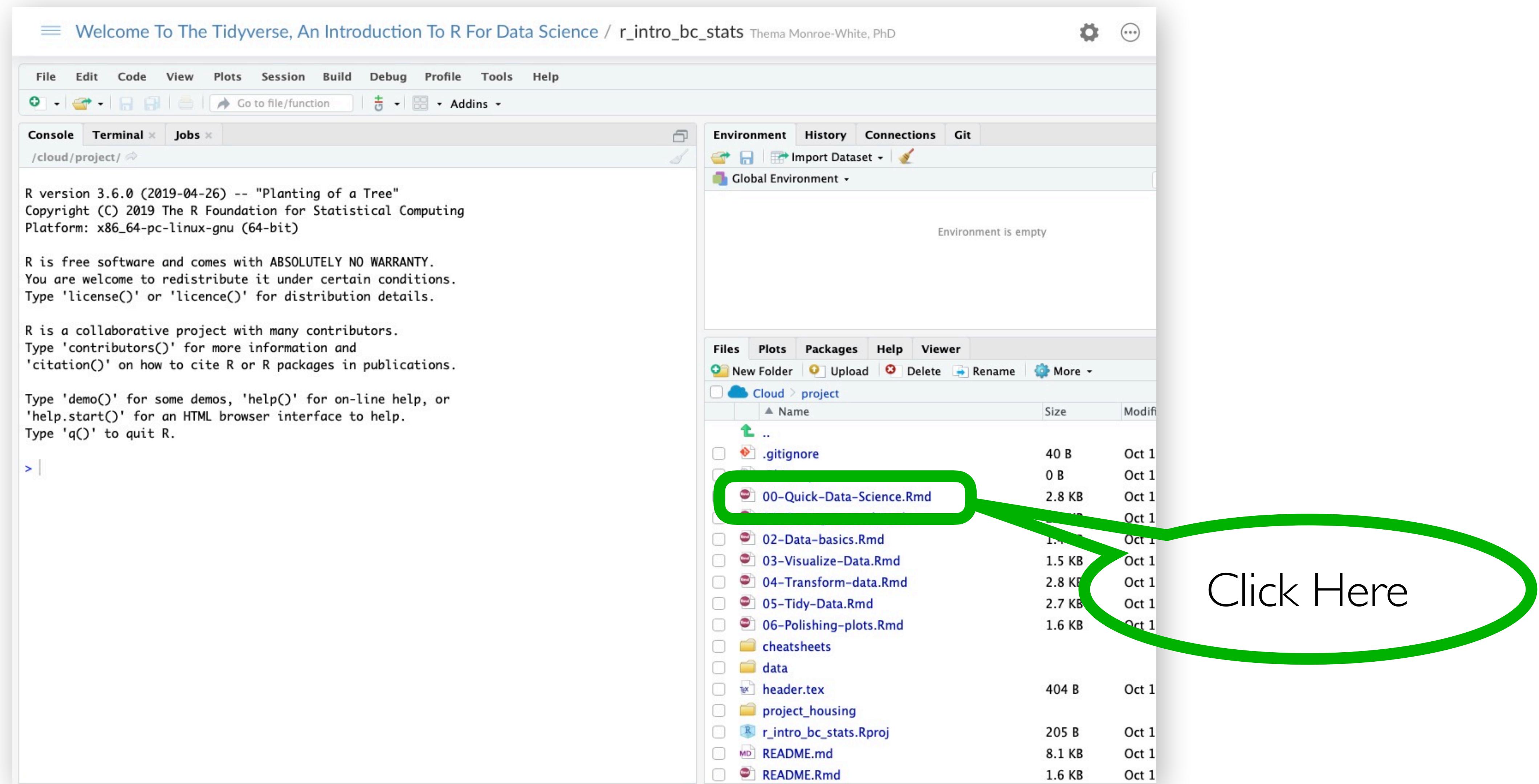
Click Here



And once you've launched the project you'll see this



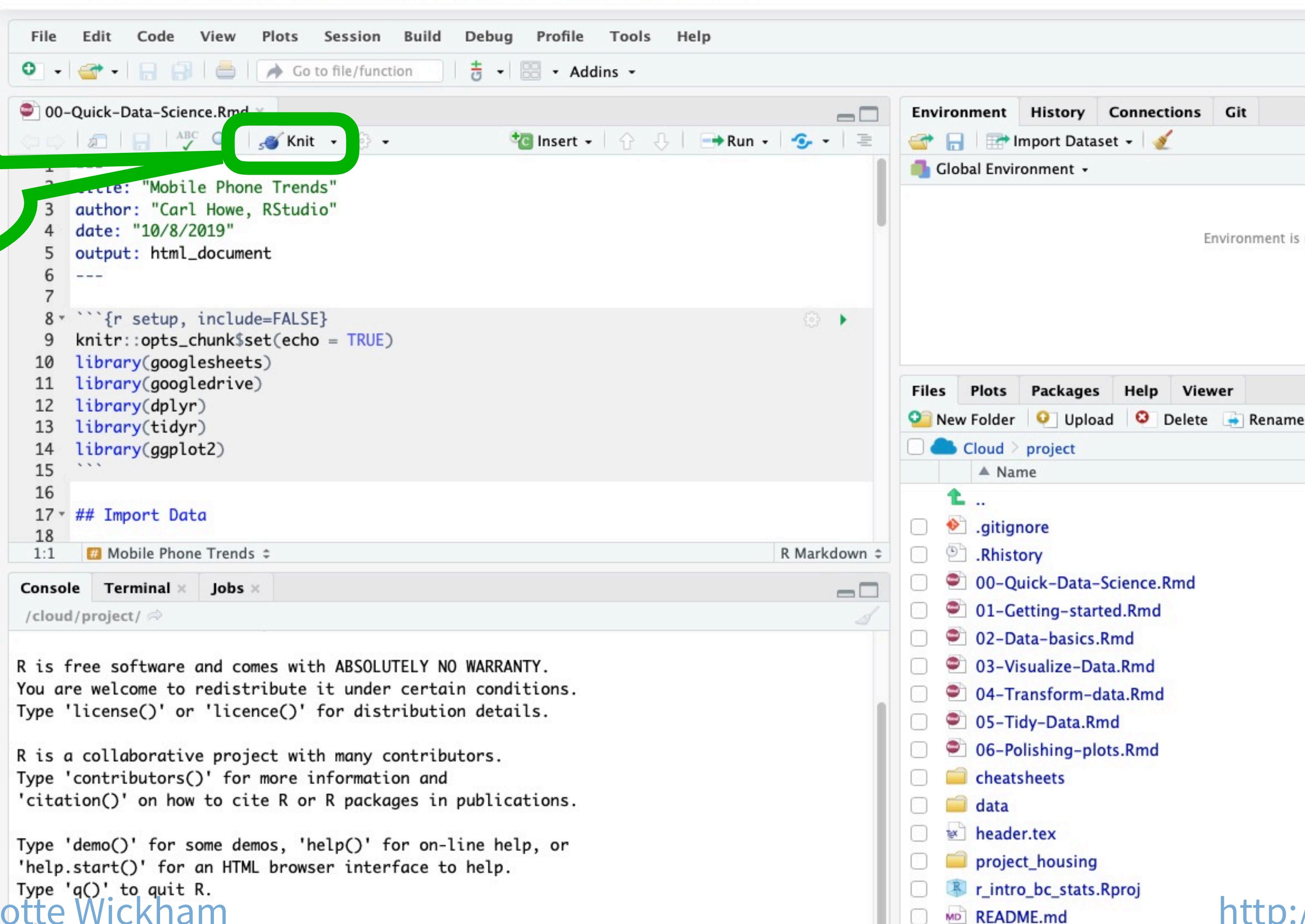
Once the project deploys, open 00-Quick-Data-Science.Rmd



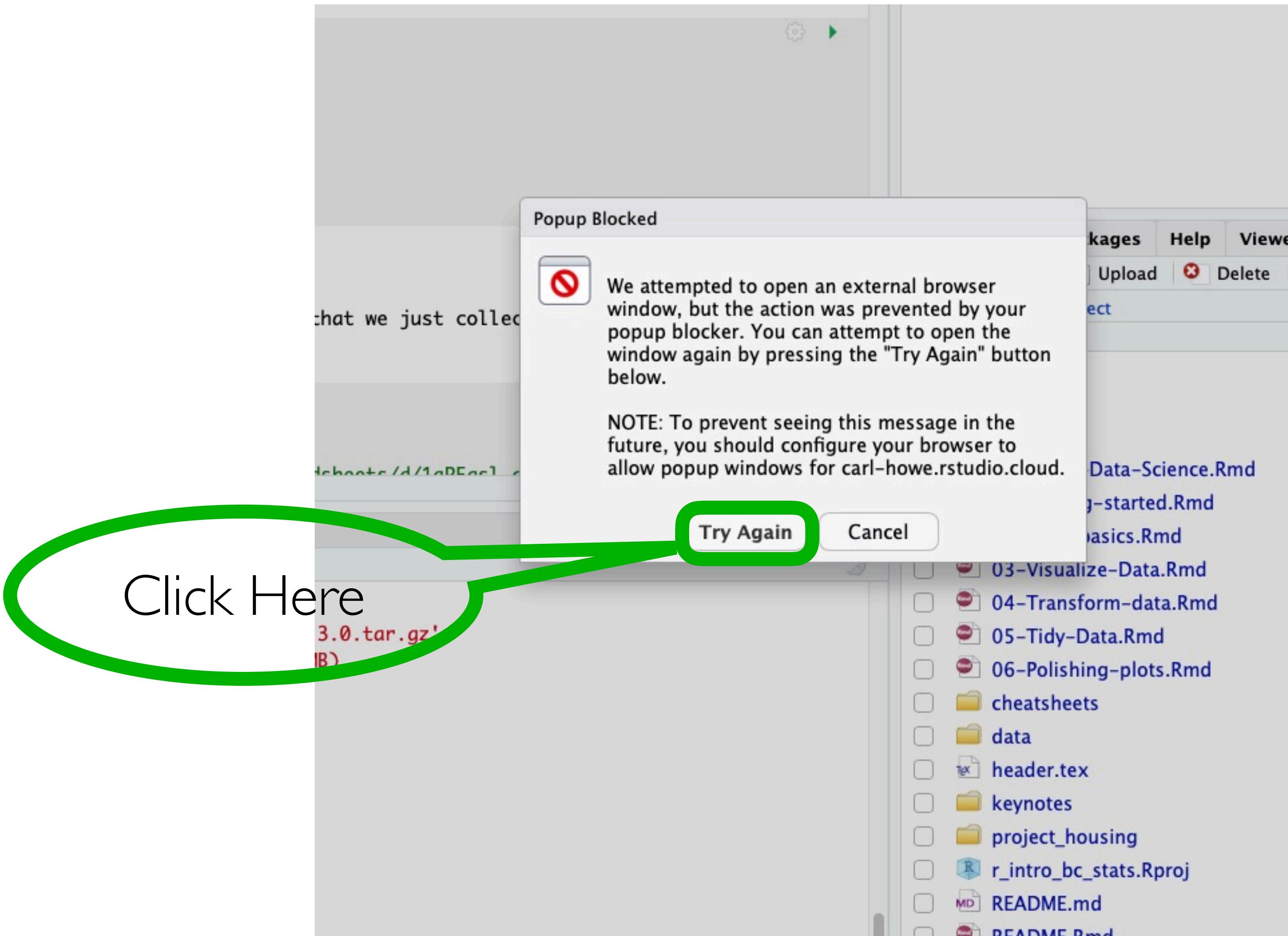
Finally, Knit your result

Welcome To The Tidyverse, An Introduction To R For Data Science / r_intro_bc_stats Thema Monroe-White, PhD

Click Here



When the pop up appears, click Try Again



You should see something like this...

Mobile Phone Trends

Carl Howe, RStudio

10/8/2019

Import Data

We can read in the Shortest Mobile Phone Survey data that we just collected using the `googlesheets` package. This will take a moment or two to run.

We now have the data in a data frame or tibble called `responses` and it has 5 survey responses in it.

Tidying the Data

Let's look at the first 5 rows of the data to get a sense of how it is formatted.

```
## # A tibble: 5 x 4
##   Timestamp    CurrentPhone  PreviousPhone  NextPhone
##   <chr>        <chr>          <chr>          <chr>
## 1 10/8/2019  13... Apple        Apple          Apple
## 2 10/8/2019  13... A manufacturer oth... Apple        I don't plan to bu...
## 3 10/8/2019  13... Apple        A manufacturer oth... Apple
## 4 10/8/2019  13... Apple        Apple          Apple
## 5 10/8/2019  13... A manufacturer oth... Apple        A manufacturer oth...
```

This data looks pretty tidy (we'll define what that means later), so we don't have to do anything in this step.

Data Transformation

We'll summarize the results to prepare them for visualization. We'll just count the proportion of Apple responses in each column to get a sense for

Your Turn

Do you understand the data?

What can you conclude from this analysis?

What other questions arise as the result of this analysis?

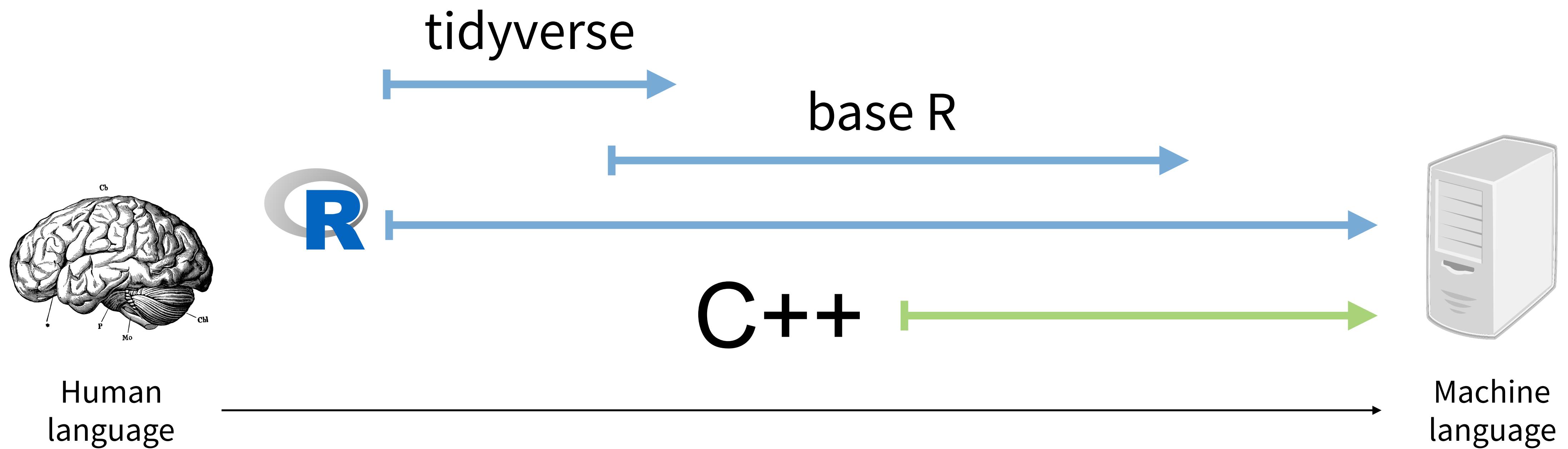
Data Science is the process
of creating insights using data

R

Why R?

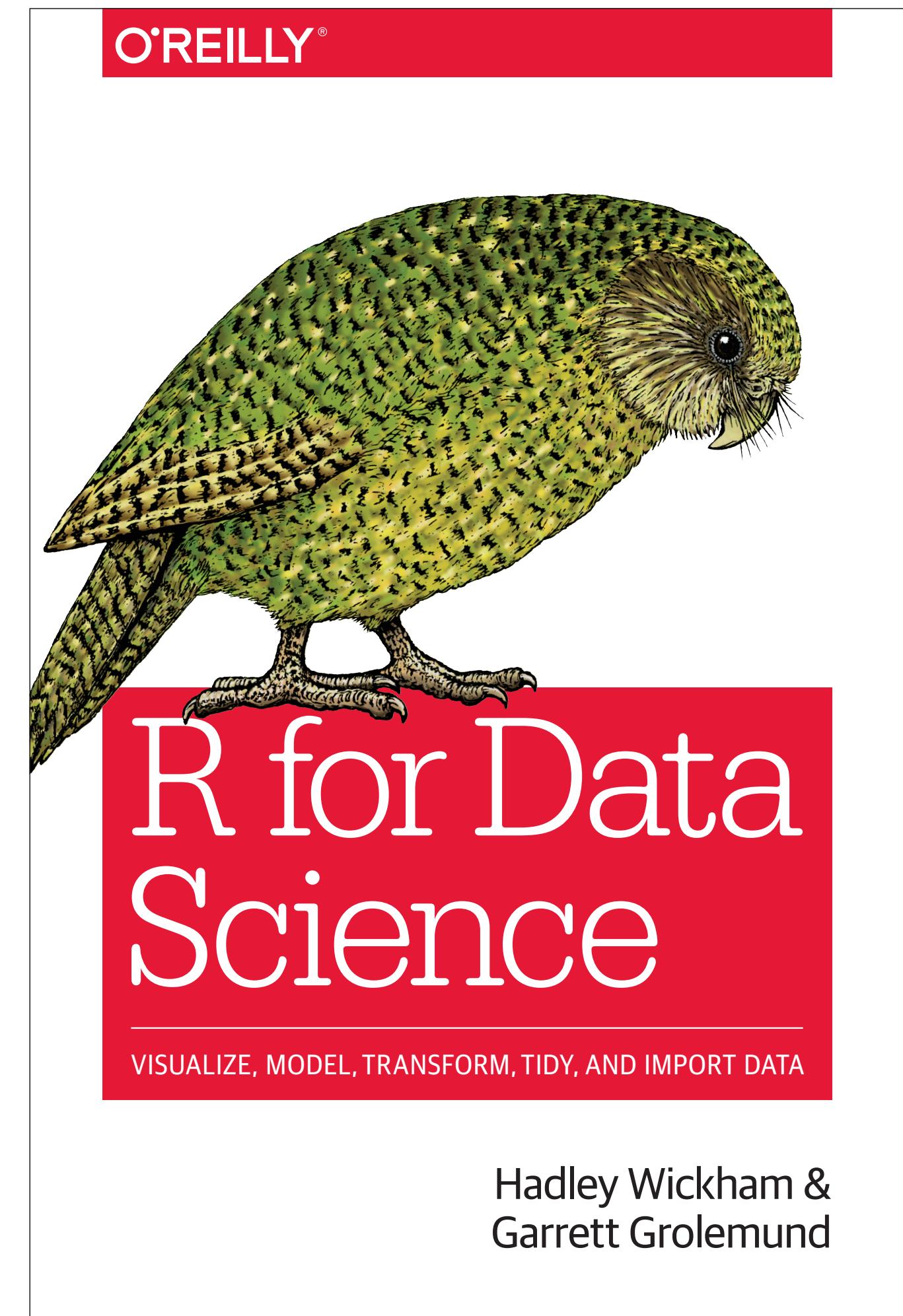


R - A computer language for scientists



Online at:

<http://r4ds.had.co.nz/>



Document

Single, 9-inch PIE CRUST

$\frac{1}{2}$ C Crisco
 $\frac{1}{2}$ t salt

Sift flour & salt into bowl. Take out $\frac{1}{4}$ C flour & set aside. Cut Crisco into remaining water till pieces are size of small peas. Mix water with flour & add paste to other mixture. Mix & shape into ball. Place in pan, flute edges, prick sides to fit 9-in. pan. Roll on floured board and bottom generously with fork, and bake in hot oven (450) for 12 to 15 min.

$1\frac{1}{4}$ C sifted flour
3 T water

Pastry for:
9 INCH PIE

LATTICE CRUST 9 INCH
1/3
2/3 C Crisco
2 C sifted flour
1 t salt
 $\frac{1}{4}$ C water

salt into bowl. Take out $\frac{1}{3}$ Crisco into remaining flour until the size of small peas. Mix $\frac{1}{4}$ C flour to form paste. Add flour paste to flour mixture. Mix & shape into a ball. Dough into 2 parts and proceed to make crusts. Use $\frac{1}{2}$ of the lattice crust recipe for a lattice pie. (in 1/3)

450° oven - 12-15 min.



Communicate

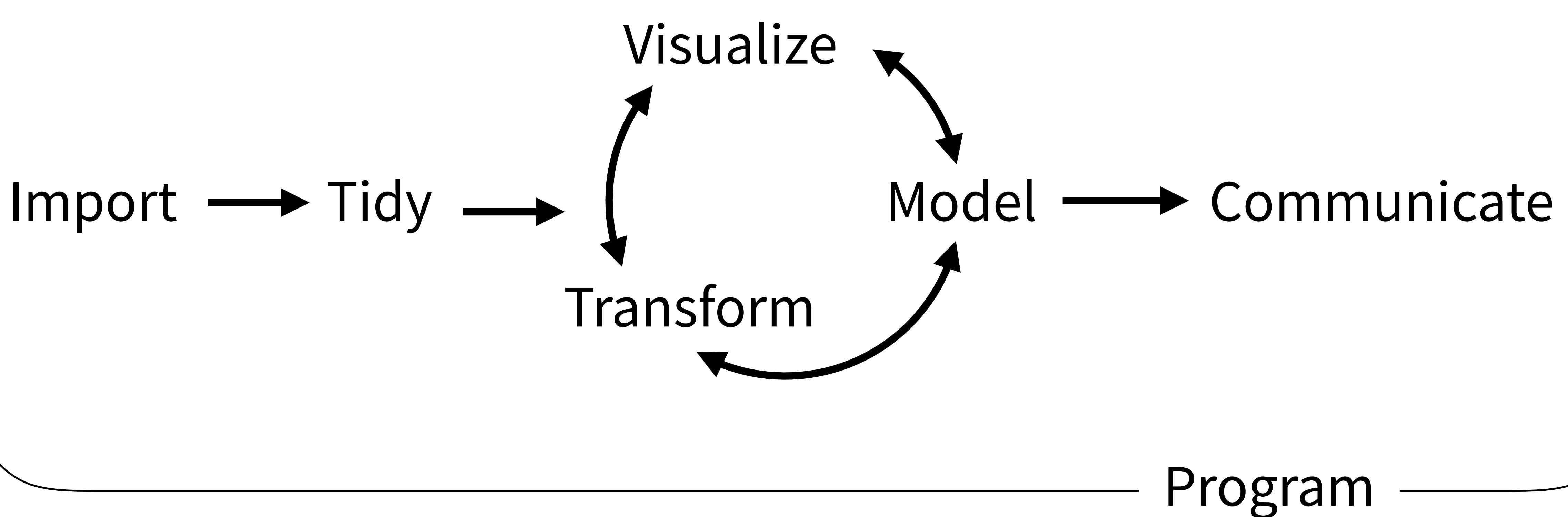
Image credit: Poster Session widdowquinn <https://flic.kr/p/7KD1iU> by cbna

Automate

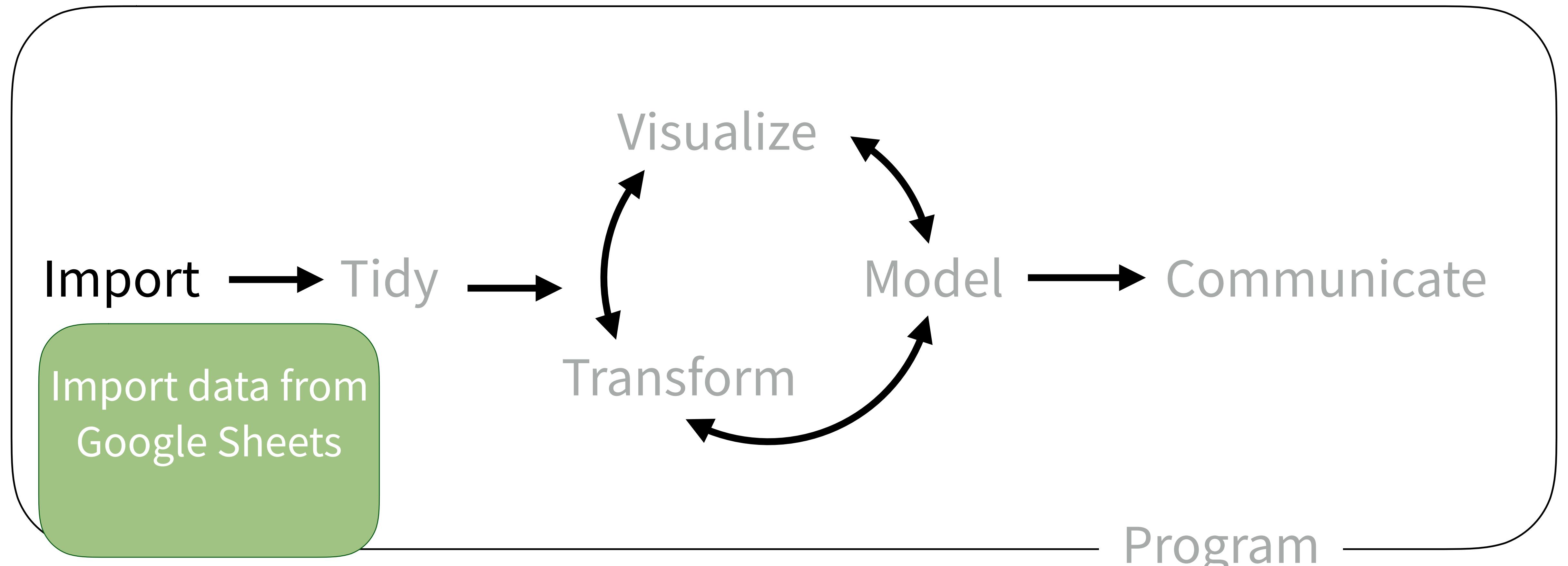
Image credit: Rover 200 Framing Line Spencer Cooper <https://flic.kr/p/cp5WgS> by cbd



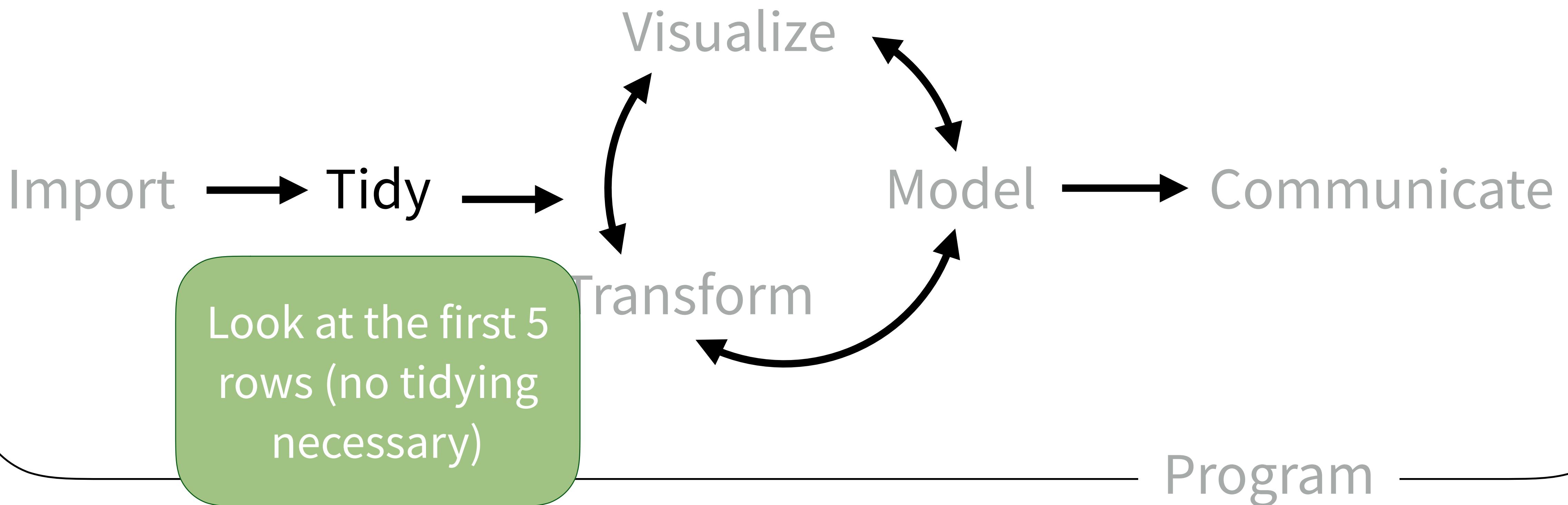
(Applied) Data Science



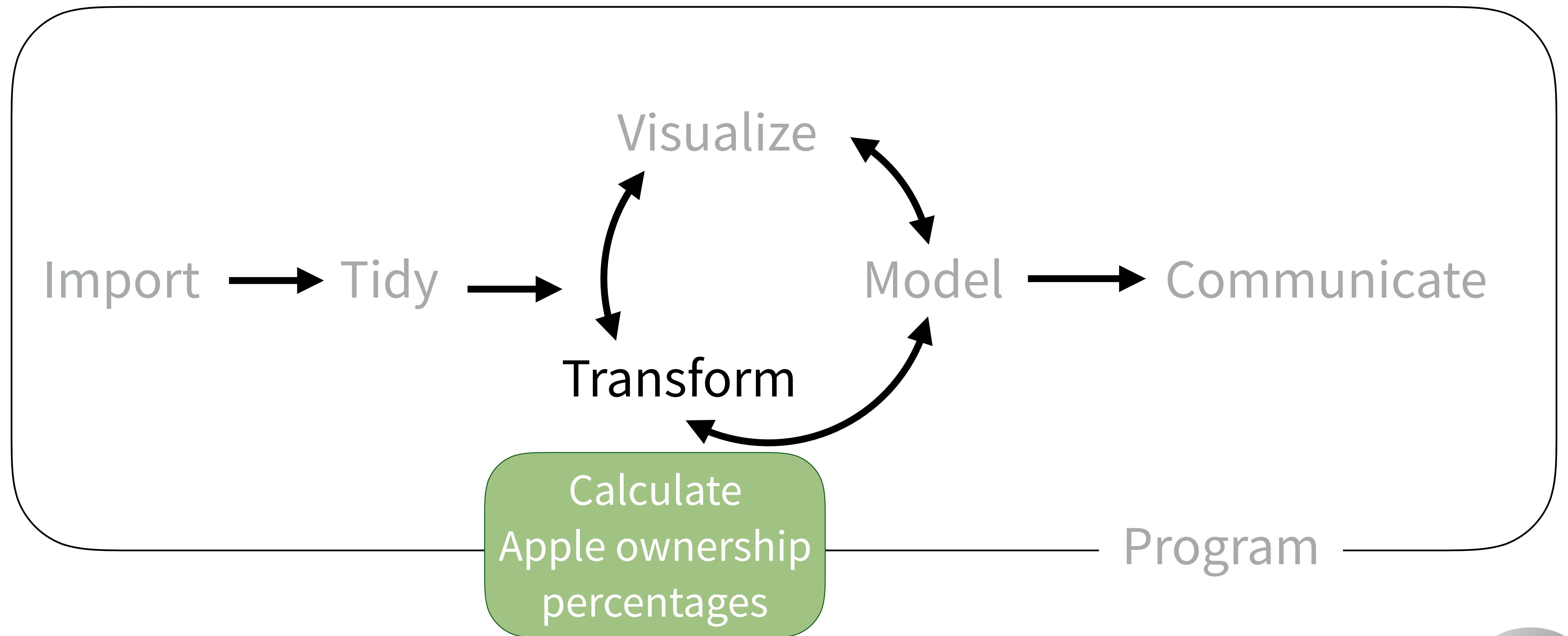
(Applied) Data Science



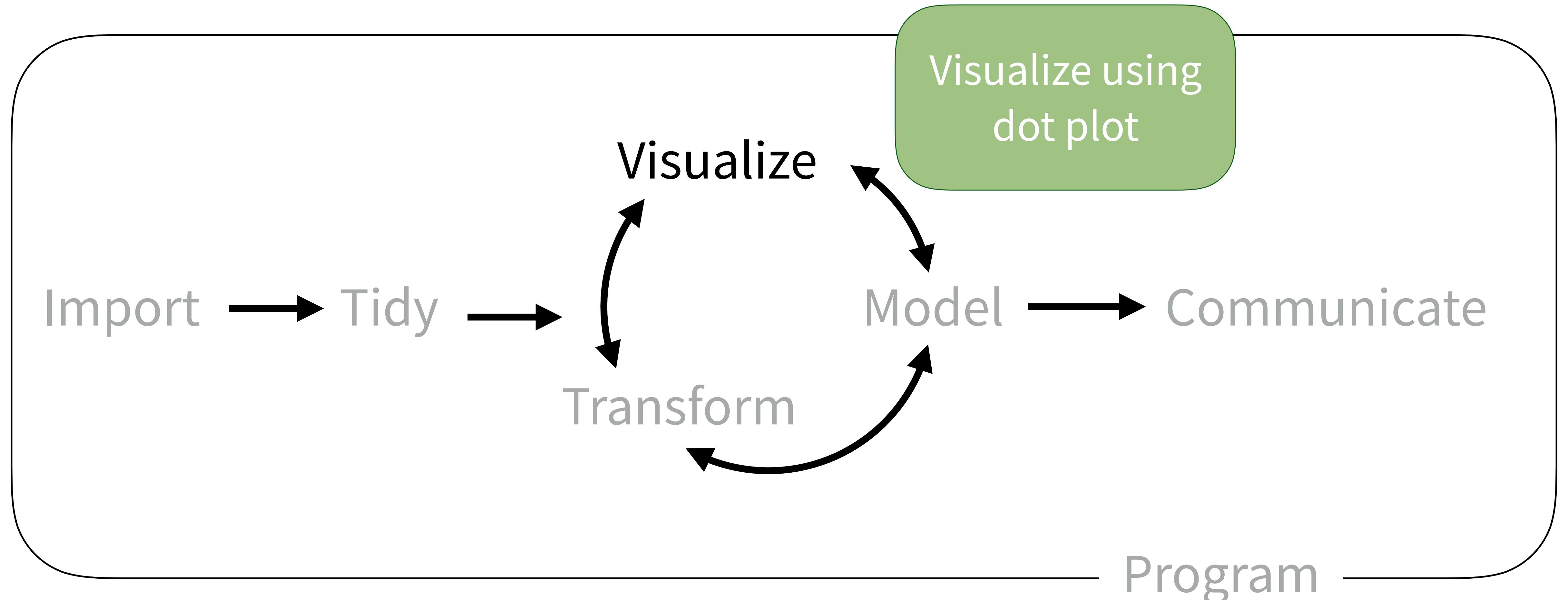
(Applied) Data Science



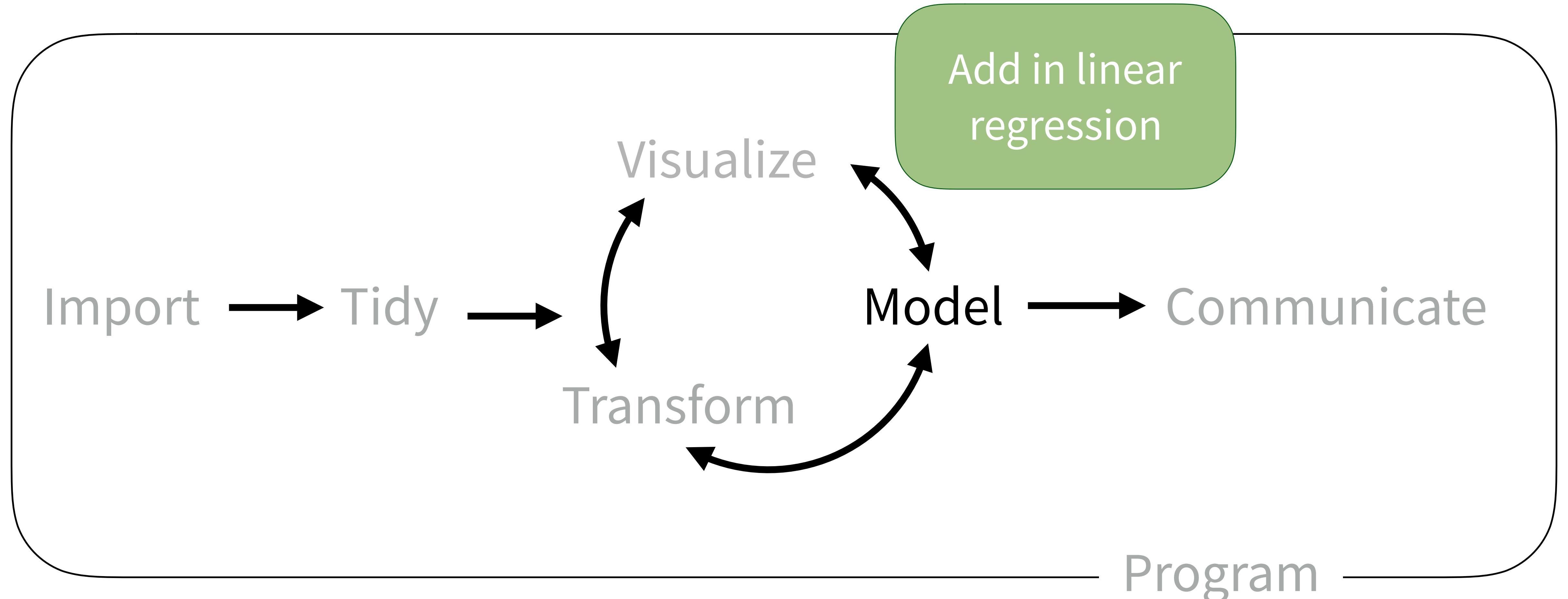
(Applied) Data Science



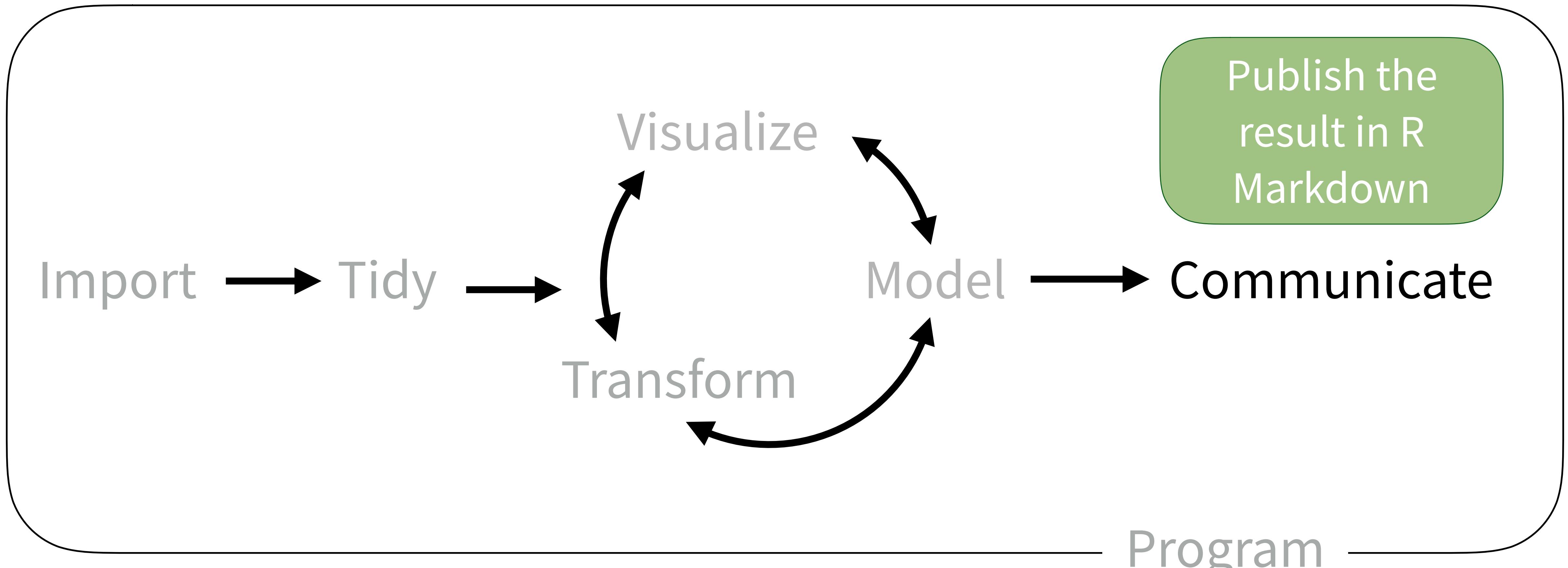
(Applied) Data Science



(Applied) Data Science



(Applied) Data Science



Summary

In this quick introduction, we've learned

- What data science is
- Why we're teaching data science with R
- A process for creating data science results

Next, we'll dive into the Tidyverse