R For Data Science Book Club – Ch. 1 & 2

Luke Morris

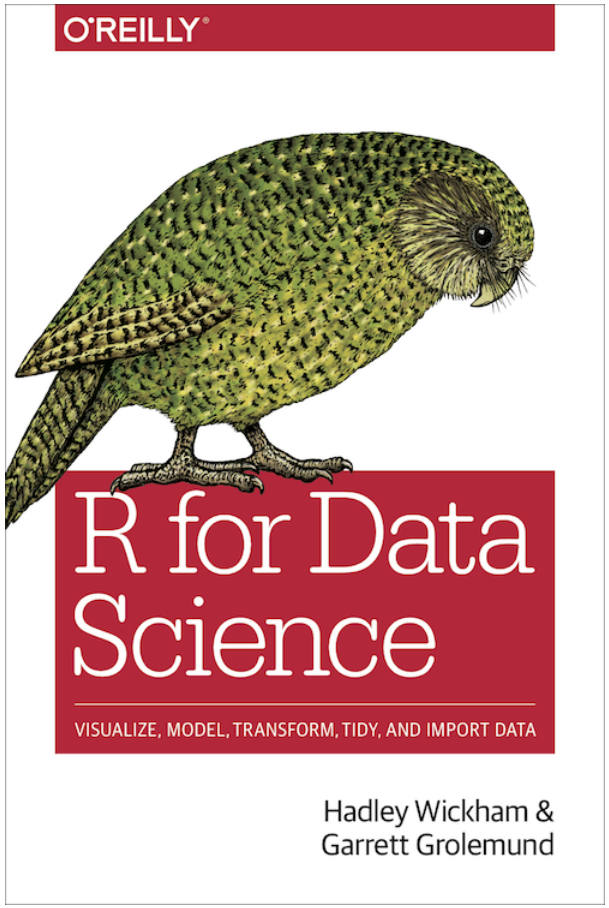
7/25/2020

Welcome!

We'll be covering a different chapter of R for Data Science by Hadley Wickham & Garrett Grolemund every Friday at 8 p.m. Eastern/5 p.m. Pacific.

Sessions will be recorded if you can't make it live every week.

Are you up for presenting one week? Let me know!



What we'll learn

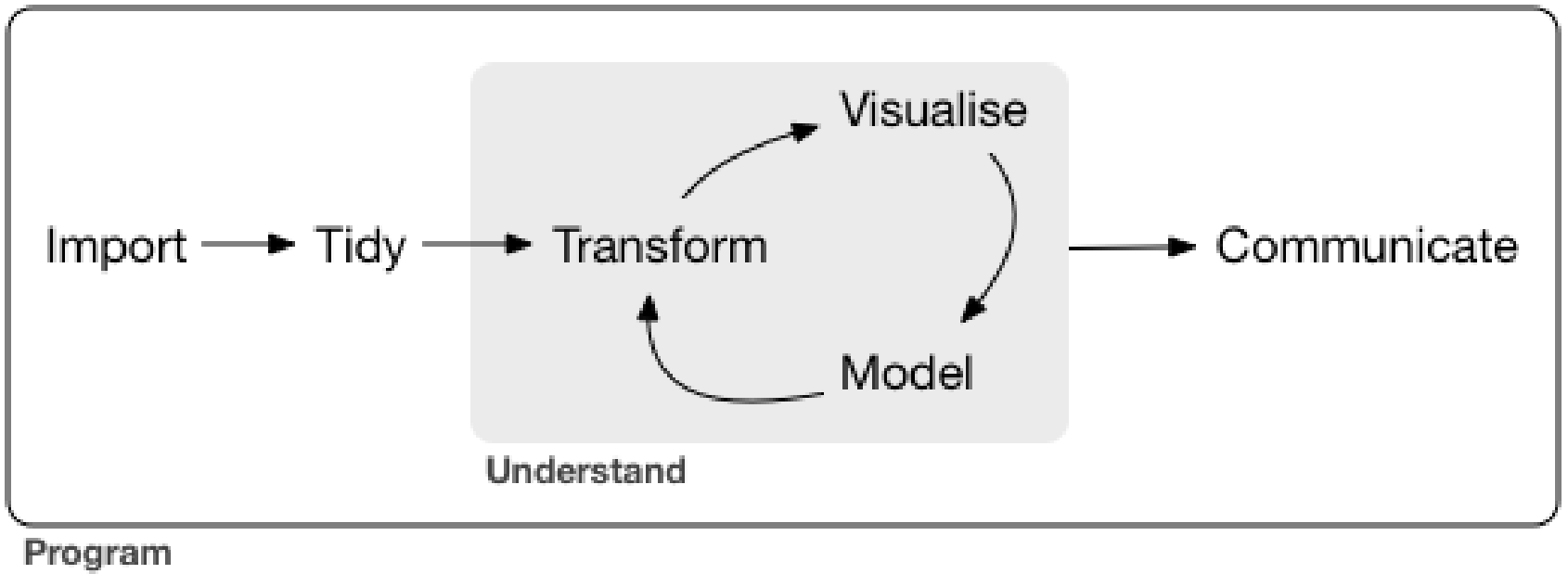
"You'll learn how to get your data into R, get it into the most useful structure, transform it, visualize it and model it.

…

"Just as a chemist learns to clean test tubes and stock a lab, you'll learn how to clean data and draw plots — and many other things besides."

— *Wickham & Grolemund, welcome page*

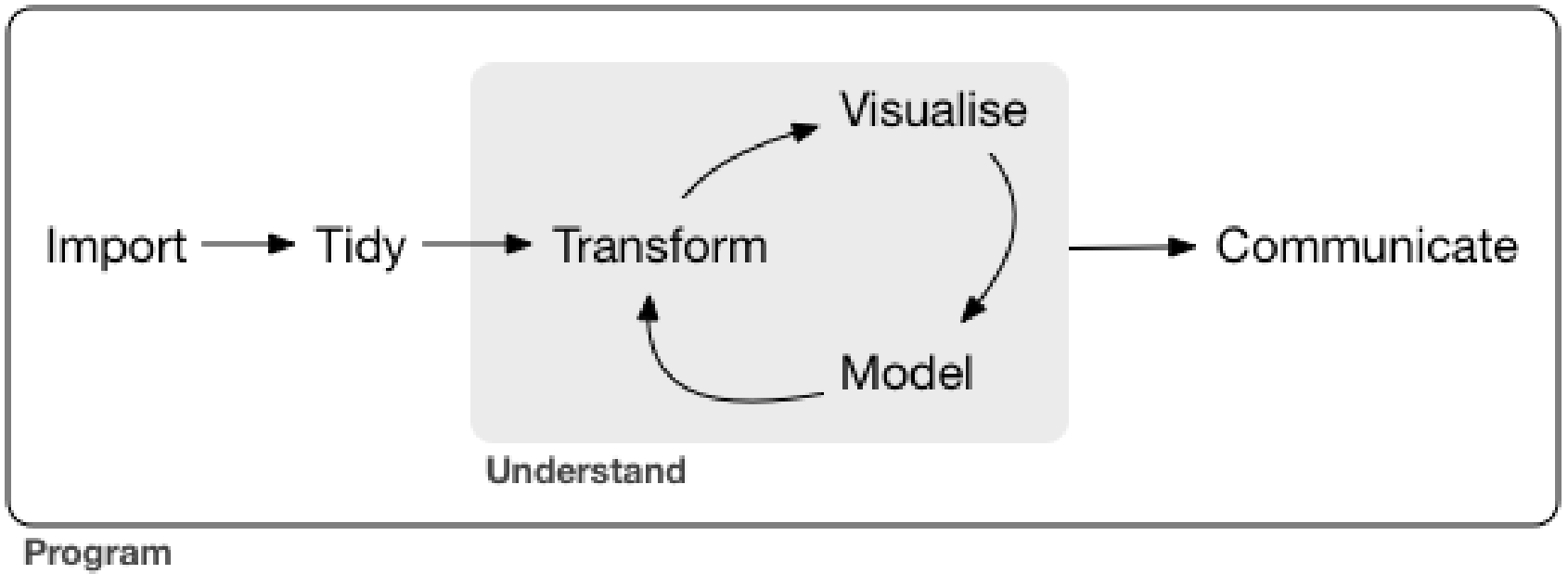
The Data Science Project Model



– Wickham & Grolemund, Ch. 1.1

Import – Getting you data into R

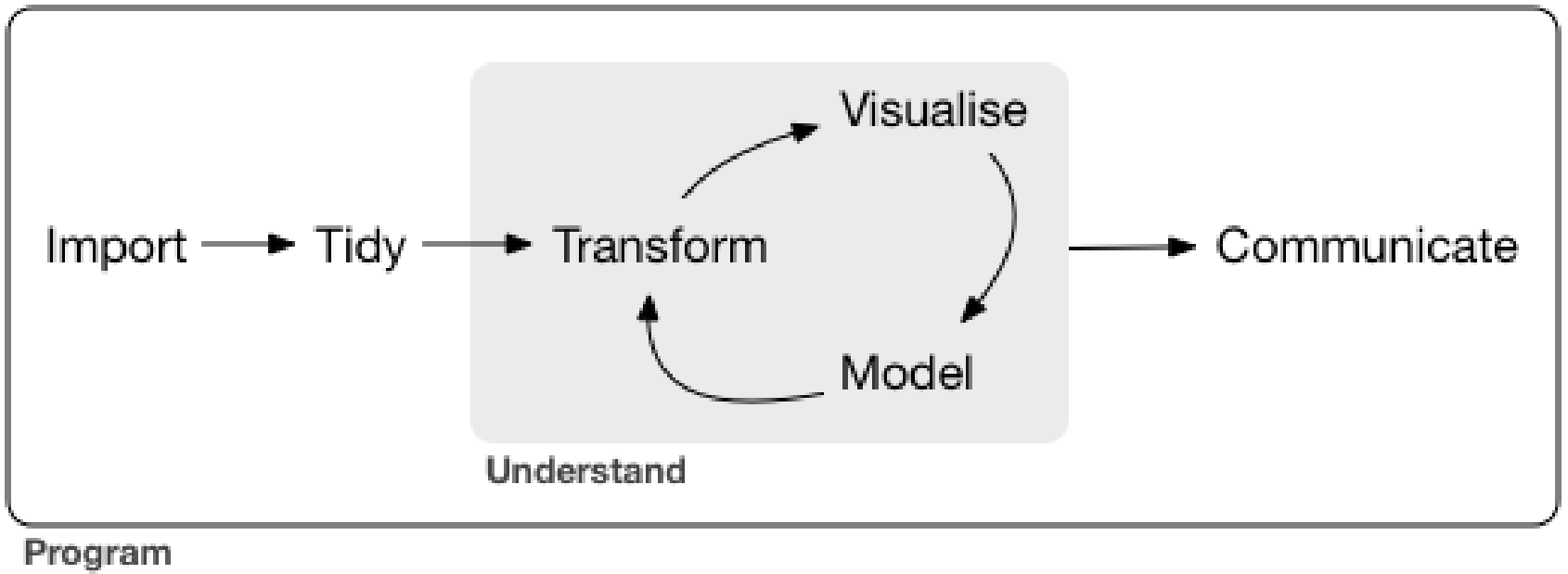
The Data Science Project Model



– Wickham & Grolemund, Ch. 1.1

Tidy – Storing your data in a consistent form

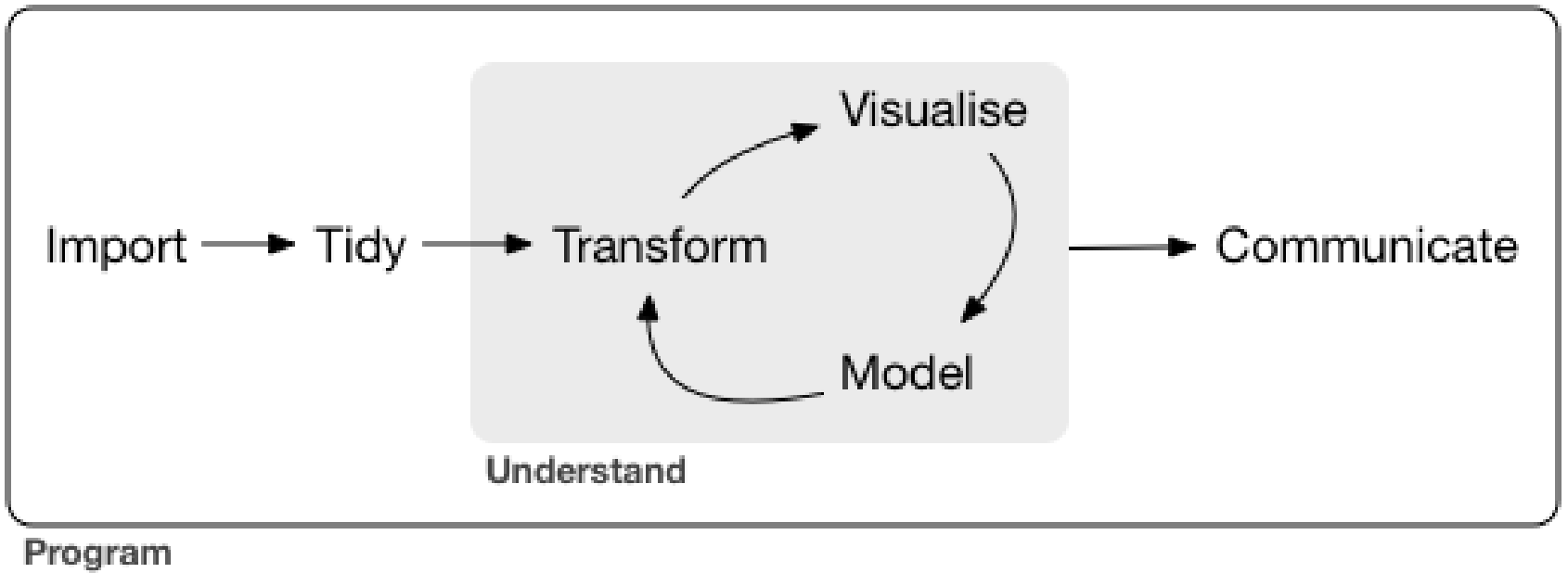
The Data Science Project Model



– Wickham & Grolemund, Ch. 1.1

Transform – Narrowing in on observations of interest and creating new variables from ones already existing

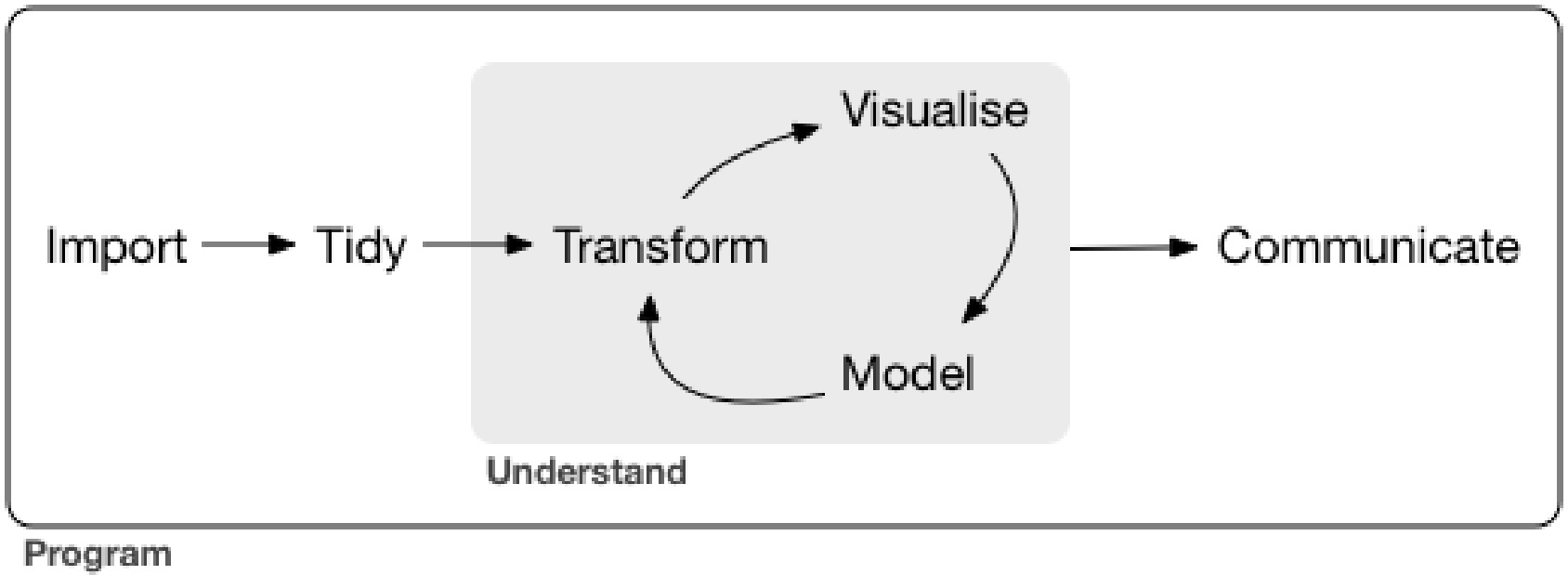
The Data Science Project Model



– Wickham & Grolemund, Ch. 1.1

Visualize – Visually representing your data

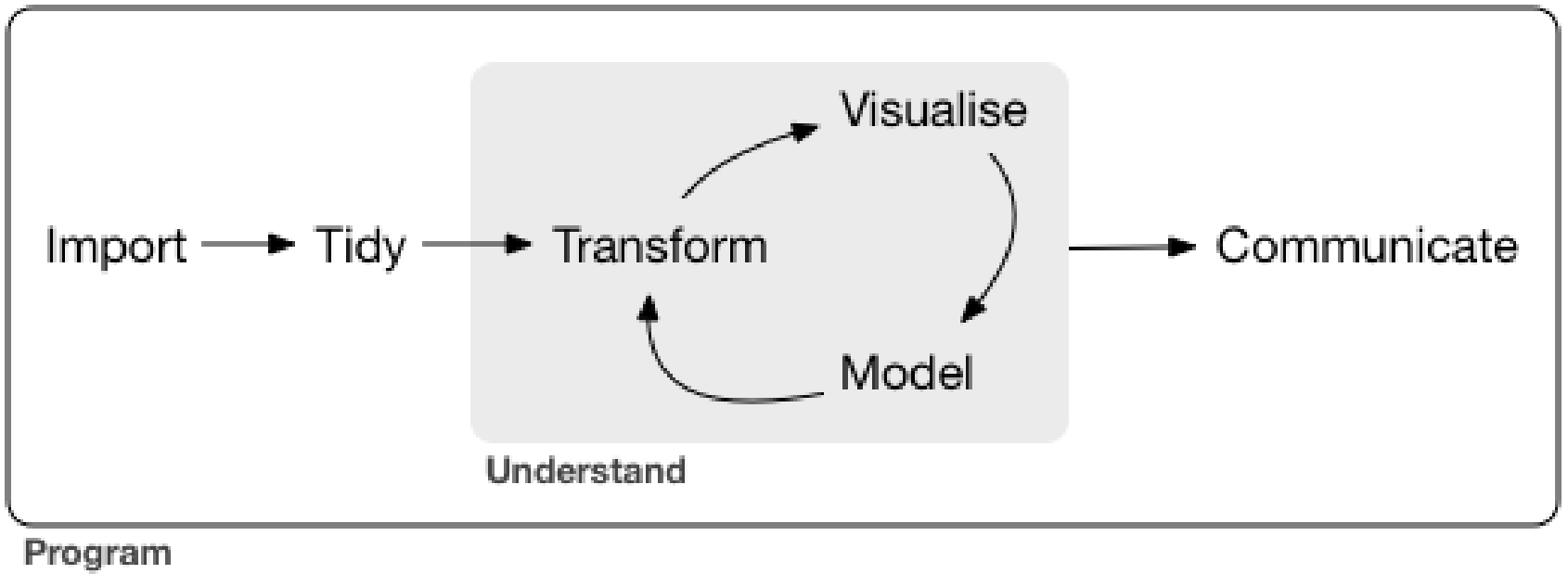
The Data Science Project Model



– Wickham & Grolemund, Ch. 1.1

Model - Using your data to answer the question at hand

The Data Science Project Model



– Wickham & Grolemund, Ch. 1.1

Communicate - Relaying your findings to others

“You don’t need to be an expert programmer to be a data scientist, but becoming a better programmer allows you to automate common tasks, and solve new problems with greater ease.”

– Wickham & Grolemund, Ch. 1.1

The 80-20 Rule

"You can tackle about 80% of every project using the tools that you'll learn in this book, but you'll need other tools to tackle the remaining 20%."

– Wickham & Grolemund, Ch. 1.1

That final 20% will vary by project, the data you're exploring and the questions you're hoping to answer.

A Wise Word from the Authors

“While it’s tempting to skip the exercises, there’s no better way to learn than practicing on real problems.”

Fortunately, there’s someone who’s run through this book before and created a sort of answer key.

A Wise Word from the Authors

“While it’s tempting to skip the exercises, there’s no better way to learn than practicing on real problems.”

Fortunately, there’s someone who’s run through this book before and created a sort of answer key.

(But seriously … you’ll learn more if you at least try to solve it on your own first)

Run into trouble? You're not alone

Confused by an error you received?

Try Googling the error message.

Other places to look for help:

Stack Overflow

Community.Rstudio (Imagine a less-grouchy Stack Overflow)

R4DS Slack

What You Need for this Book Club

1. Install R
2. Install RStudio
3. Open RStudio and install the Tidyverse package by typing or pasting the following in the console (bottom-left box)

```
install.packages("tidyverse")
```

1. Install these other packages the book uses as examples by typing or pasting the following in the console

```
install.packages(c("nycflights13", "gapminder", "Lahman"))
```

See you next week!

We'll be tackling **Data Visualization**