Review & Best practices

DAY 2, SESSION 5

September 18, 2023

Review (continued)!

Welcome, **Brendan!**

We're glad you're here, because it is more fun to learn data science together.

HELPFUL LINKS

Content Summary Tips for presentin... Demo milestone...

NEXT LESSON:

Welcome to R

Continue

LEARNING PATH

Full View Student View

Week 6
3 optional items
MILESTONE 6
Conf Session 1 - Review

Babynames Exploration (Cas...
Optional

Joining and exploring flights...
Optional

Practice cleaning data
Optional

Week 7
4 optional items
MILESTONE 7
Conf Session 2 - EDA

Week:

1 2 3 4 5 6 7 8 9 10 11 12

Best practices

Welcome, **Brendan!**

We're glad you're here, because it is more fun to learn data science together.

HELPFUL LINKS

Content Summary Tips for presentin... Demo milestone...

NEXT LESSON:

Welcome to R

Continue

LEARNING PATH

Full View Student View

Week 10

MILESTONE 10
Conf Session 5 - Best practices

Week 11

MILESTONE 11
Conf Project - Employee Attrition

Week 12
5 optional items

MILESTONE 12
Conf Session 7 - Data types

Types and classes
Optional

Basic operations with strings
Optional

Week:

1 2 3 4 5 6 7 8 9 10 11 12

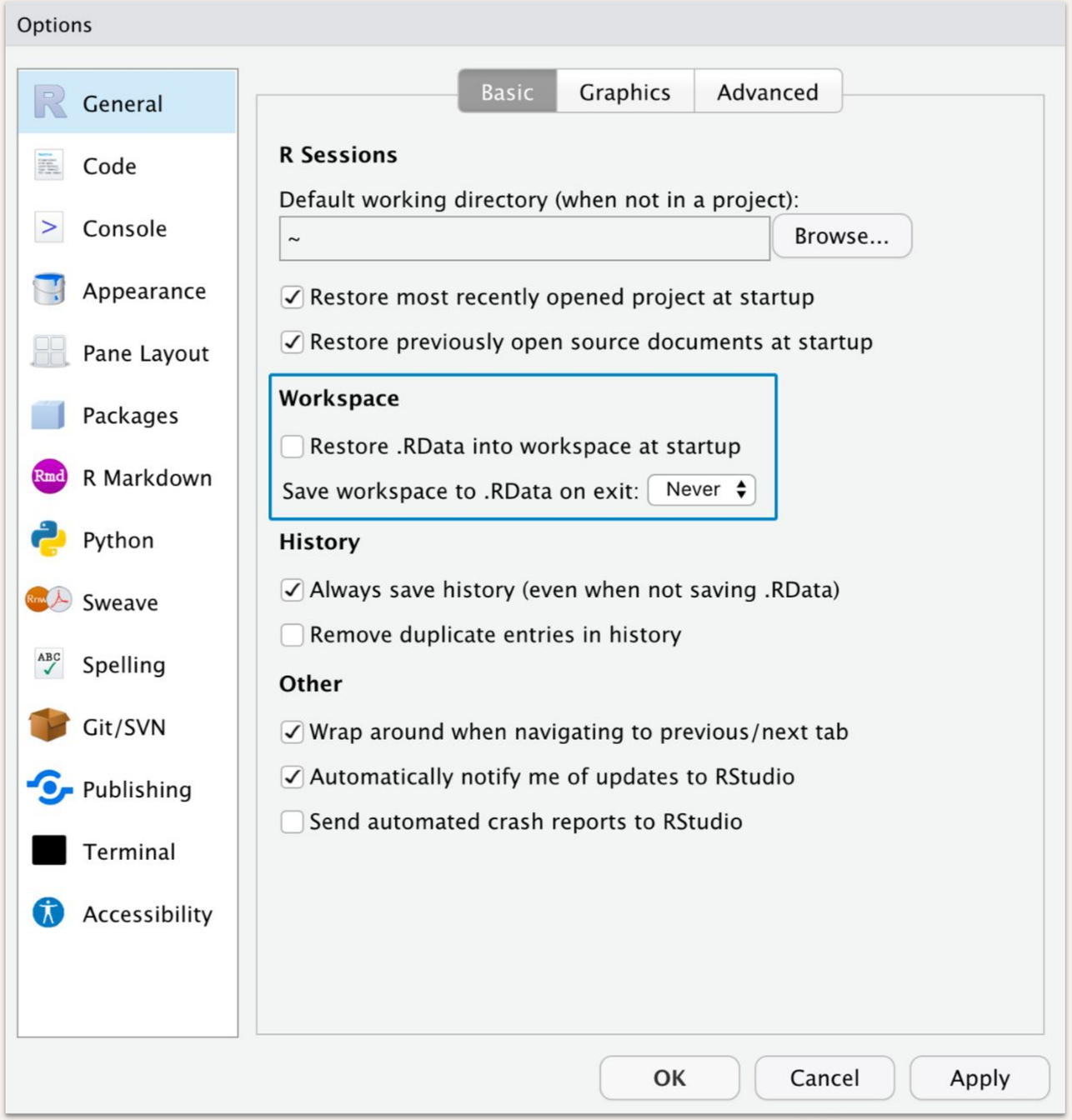
Agenda

1. Customize your workspace
2. Project-oriented workflows
3. Naming files
4. Reproducible environments
5. Code style
6. Version control
7. Asking for help

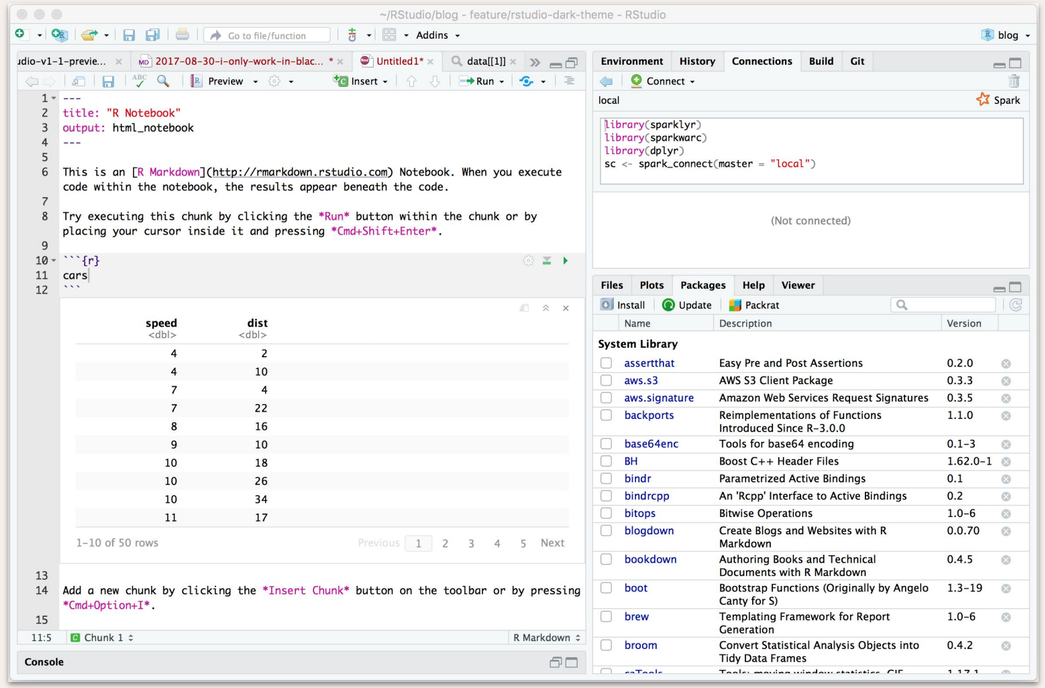
1. Customize your workspace

IDE settings

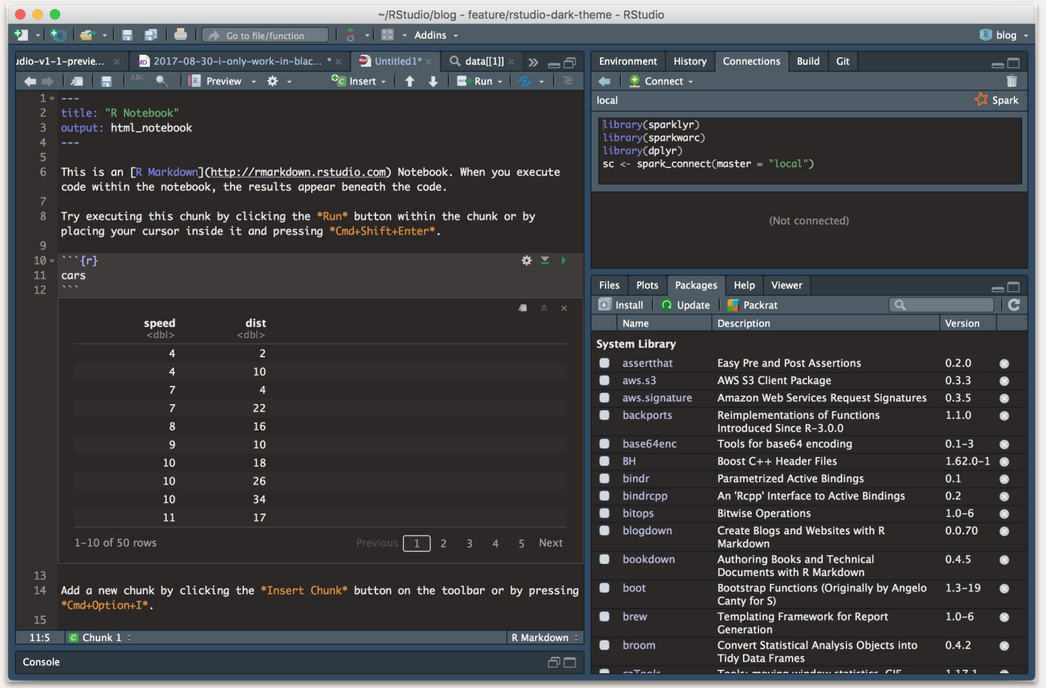
Tools > Global Setup



Themes

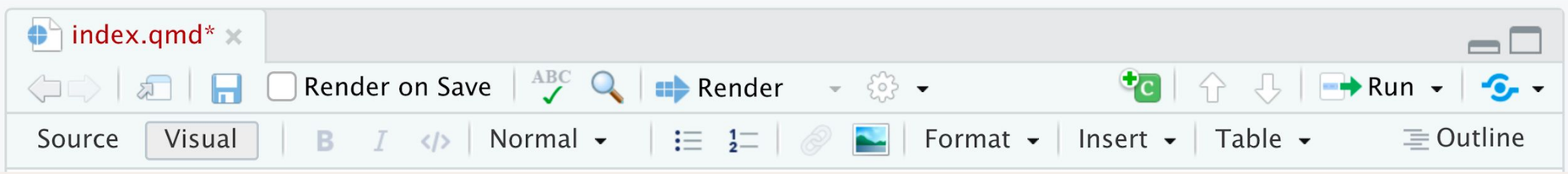

~/RStudio/blog - feature/rstudio-dark-theme - RStudio
Go to file/function
Addins
blog
2017-08-30-i-only-work-in-blac...
Untitled1*
data[[1]]
Preview
Insert
Run
Environment
History
Connections
Build
Git
Connect
local
Spark
library(sparklyr)
library(sparkwarc)
library(dplyr)
sc <- spark_connect(master = "local")
(Not connected)
Files
Plots
Packages
Help
Viewer
Install
Update
Packrat
Name
Description
Version
System Library
assertthat
Easy Pre and Post Assertions
0.2.0
aws.s3
AWS S3 Client Package
0.3.3
aws.signature
Amazon Web Services Request Signatures
0.3.5
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.1.0
base64enc
Tools for base64 encoding
0.1-3
BH
Boost C++ Header Files
1.62.0-1
bindr
Parametrized Active Bindings
0.1
bindrcpp
An 'Rcpp' Interface to Active Bindings
0.2
bitops
Bitwise Operations
1.0-6
blogdown
Create Blogs and Websites with R Markdown
0.0.70
bookdown
Authoring Books and Technical Documents with R Markdown
0.4.5
boot
Bootstrap Functions (Originally by Angelo Canty for S)
1.3-19
brew
Templating Framework for Report Generation
1.0-6
broom
Convert Statistical Analysis Objects into Tidy Data Frames
0.4.2
---
title: "R Notebook"
output: html_notebook

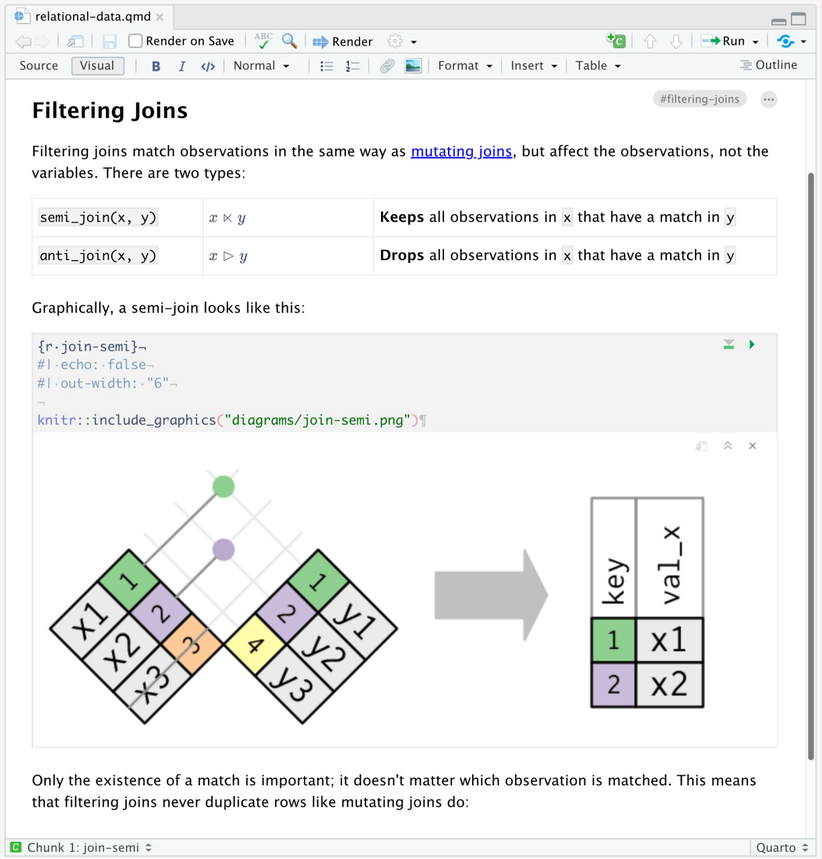
This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.
Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.
```{r}
cars
```
speed <dbl>
dist <dbl>
4 2
4 10
7 4
7 22
8 16
9 10
10 18
10 26
10 34
11 17
1-10 of 50 rows
Previous 1 2 3 4 5 Next
Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.
11:5
Chunk 1
R Markdown
Console


~/RStudio/blog - feature/rstudio-dark-theme - RStudio
Go to file/function
Addins
blog
2017-08-30-i-only-work-in-blac...
Untitled1*
data[[1]]
Preview
Insert
Run
Environment
History
Connections
Build
Git
Connect
local
Spark
library(sparklyr)
library(sparkwarc)
library(dplyr)
sc <- spark_connect(master = "local")
(Not connected)
Files
Plots
Packages
Help
Viewer
Install
Update
Packrat
Name
Description
Version
System Library
assertthat
Easy Pre and Post Assertions
0.2.0
aws.s3
AWS S3 Client Package
0.3.3
aws.signature
Amazon Web Services Request Signatures
0.3.5
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.1.0
base64enc
Tools for base64 encoding
0.1-3
BH
Boost C++ Header Files
1.62.0-1
bindr
Parametrized Active Bindings
0.1
bindrcpp
An 'Rcpp' Interface to Active Bindings
0.2
bitops
Bitwise Operations
1.0-6
blogdown
Create Blogs and Websites with R Markdown
0.0.70
bookdown
Authoring Books and Technical Documents with R Markdown
0.4.5
boot
Bootstrap Functions (Originally by Angelo Canty for S)
1.3-19
brew
Templating Framework for Report Generation
1.0-6
broom
Convert Statistical Analysis Objects into Tidy Data Frames
0.4.2
---
title: "R Notebook"
output: html_notebook

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.
Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.
```{r}
cars
```
speed <dbl>
dist <dbl>
4 2
4 10
7 4
7 22
8 16
9 10
10 18
10 26
10 34
11 17
1-10 of 50 rows
Previous 1 2 3 4 5 Next
Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.
11:5
Chunk 1
R Markdown
Console

Visual editor





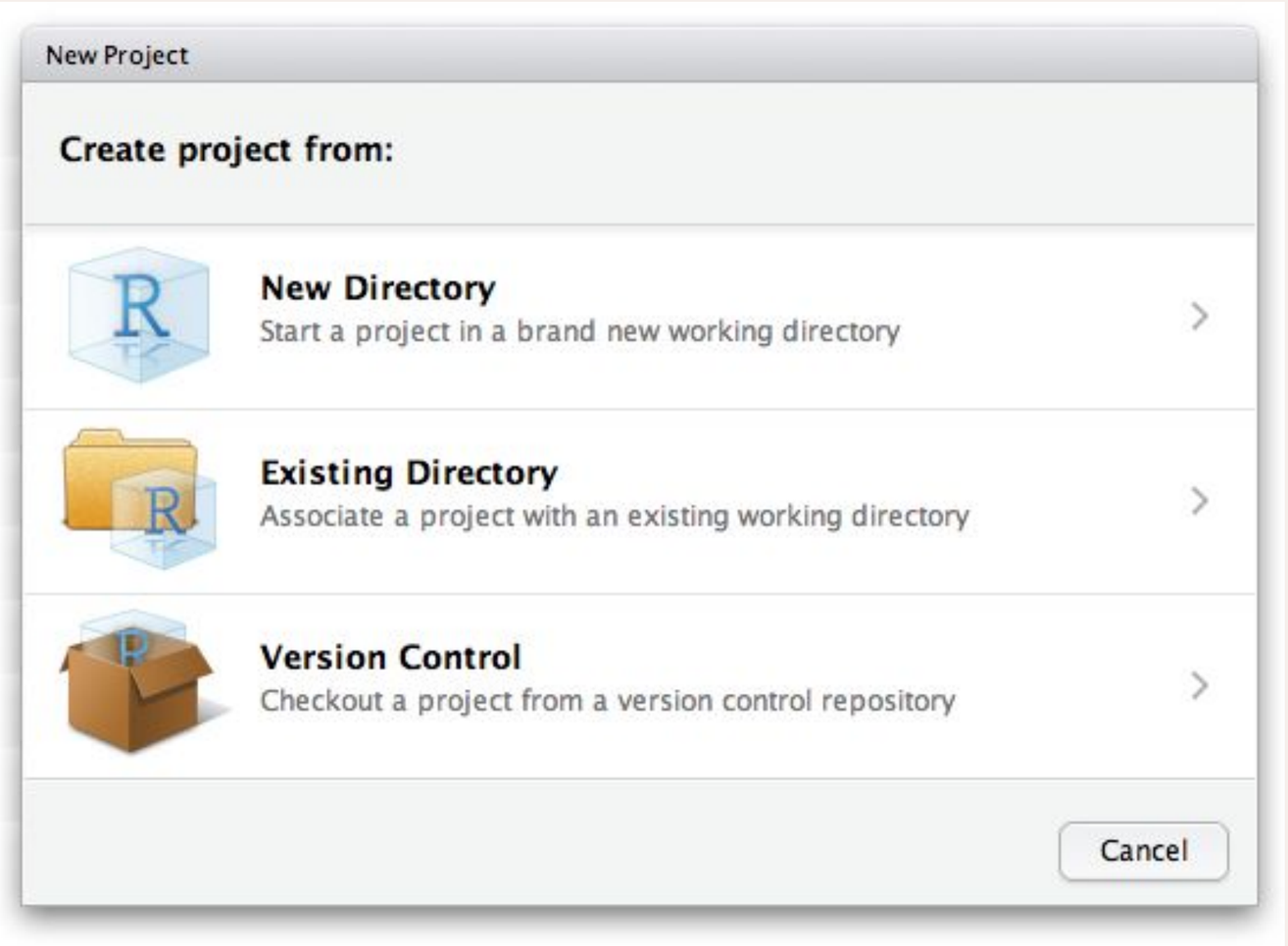
Resources

- Customizing the RStudio IDE
- Using themes in the IDE
- Rainbow parentheses
- Visual editing in RStudio

2. Project-oriented workflows

Projects

Use R projects to make your code more organized and shareable



here

Use the here package to manage file paths

❌ `setwd("path/that/only/exists/on/my/computer")`

Resources

- What They Forgot to Teach You About R Chapter 3: Project-oriented workflow
- R4DS Chapter 7: Script and Projects
- The here package

3. Naming files

Naming files

- Machine readable
- Human readable
- Sorts/orders nicely

NO

myabstract.docx
Joe's Filenames Use Spaces and Punctuation.xlsx
figure 1.png
fig 2.png
JW7d^(2sl@deletethisandyourcareerisoverWx2*.txt

YES

2014-06-08_abstract-for-sla.docx
joes-filenames-are-getting-better.xlsx
fig01_scatterplot-talk-length-vs-interest.png
fig02_histogram-talk-attendance.png
1986-01-28_raw-data-from-challenger-o-rings.txt

Image from Jenny Bryan

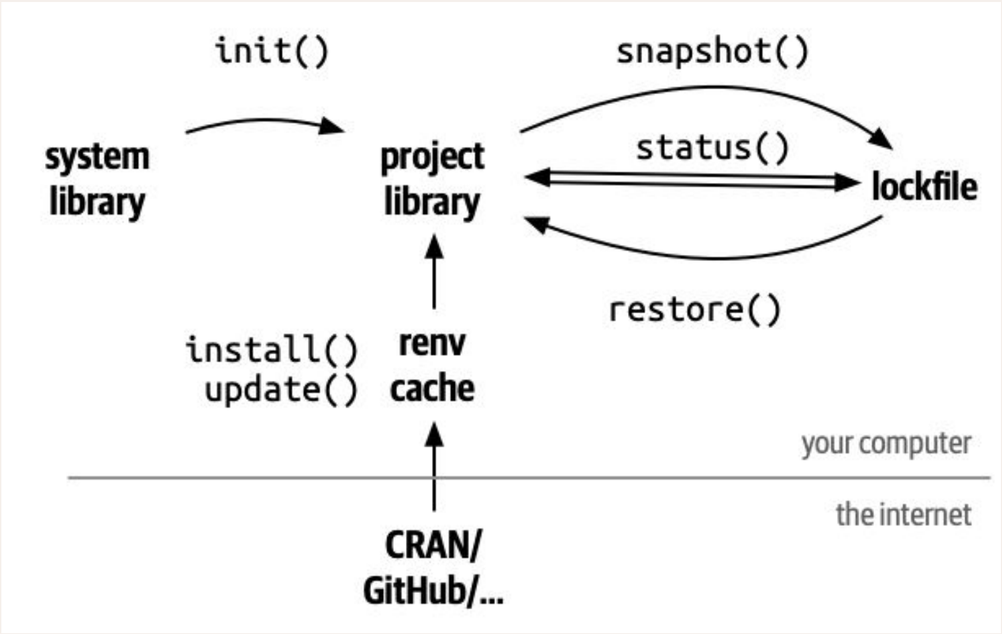
Resources

- Naming things (slides)
- Tidyverse Style Guide Chapter 1: Files

4. Reproducible environments

renv

Create reproducible environments for your R projects with the renv package



Resources

- The renv package
- You should be using renv (rstudio::conf(2022) talk)

5. Code style

Code style

Consistent style makes your easier to read & debug by your future self and collaborators

```
# Good
iris %>%
  group_by(Species) %>%
  summarize_if(is.numeric, mean) %>%
  ungroup() %>%
  gather(measure, value, -Species) %>%
  arrange(value)

# Bad
iris %>% group_by(Species) %>% summarize_all(mean) %>%
ungroup %>% gather(measure, value, -Species) %>%
arrange(value)
```

Image from the Tidyverse Style Guide

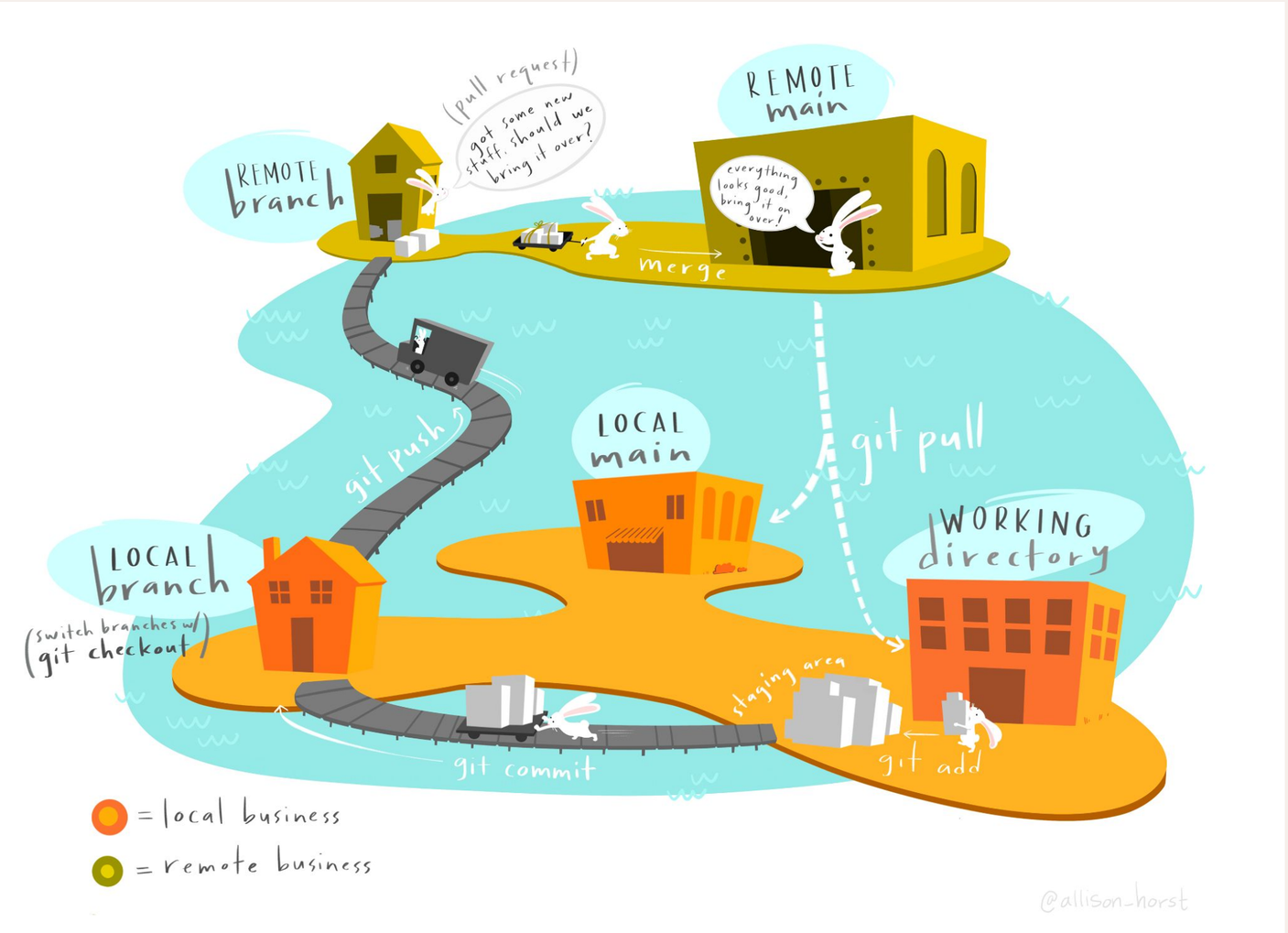
Resources

- Tidyverse style guide
- The styler package
- R4DS Chapter 5: Code Style (short primer)

6. Version control

Git & Github

Use version control to share and collaborate on code projects



Artwork by Allison Horst: https://allisonhorst.com/git-github

Resources

- Happy Git With R

7. Asking for help

Tips for searching

- Google the error message!
- Add “R” to your query
- Add a package name (e.g. “dplyr”) to narrow your results
- Try Posit Community and Stack Overflow

Reprex

Create a **repr**oducible **ex**ample when asking for help

Resources

- R4DS Chapter 9: Getting Help
- The reprex package

Questions?