



# Getting Started with R – Part II

Who am I?

---

# About Perry

---

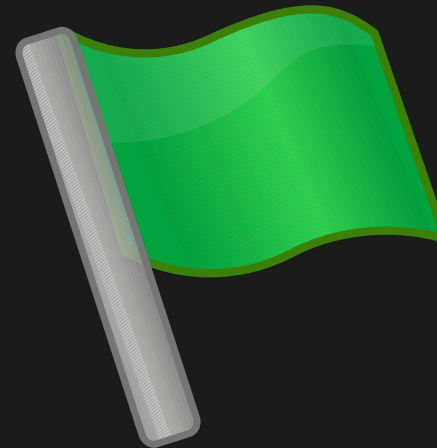
- Motivates, mobilizes, and connects cross-functional teams with technical solutions and support
- Provides customer-focused Computer Professional services with Data Science / Systems Engineering experience in commercial and non-profit industries.
- Delivers system, network, and security support in a wide variety of business and home environments.
- Partners with clients for training and end-developer support efforts, especially in the areas of configuration management, operating system integration.

In Our Last Episode...

---

# Project Overview (Revised)

- Download R
- Download Rstudio
  - Download tidyverse package
- Compile everything.
- Download Docker
- docker run *something something*
- Done!

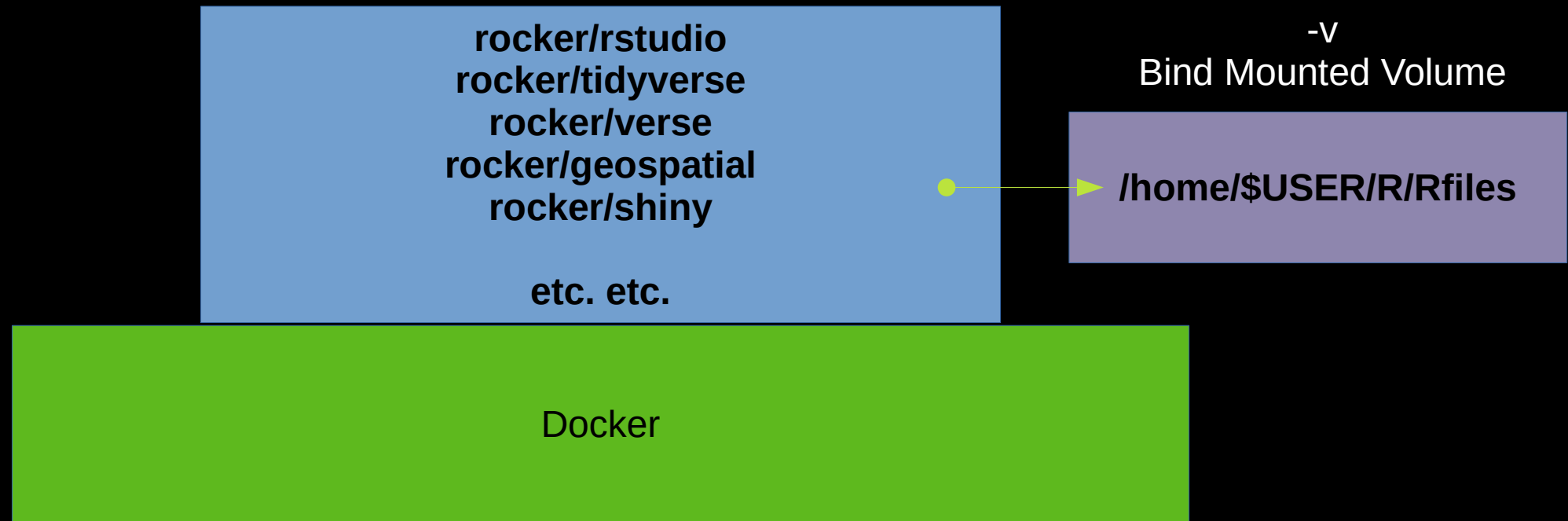




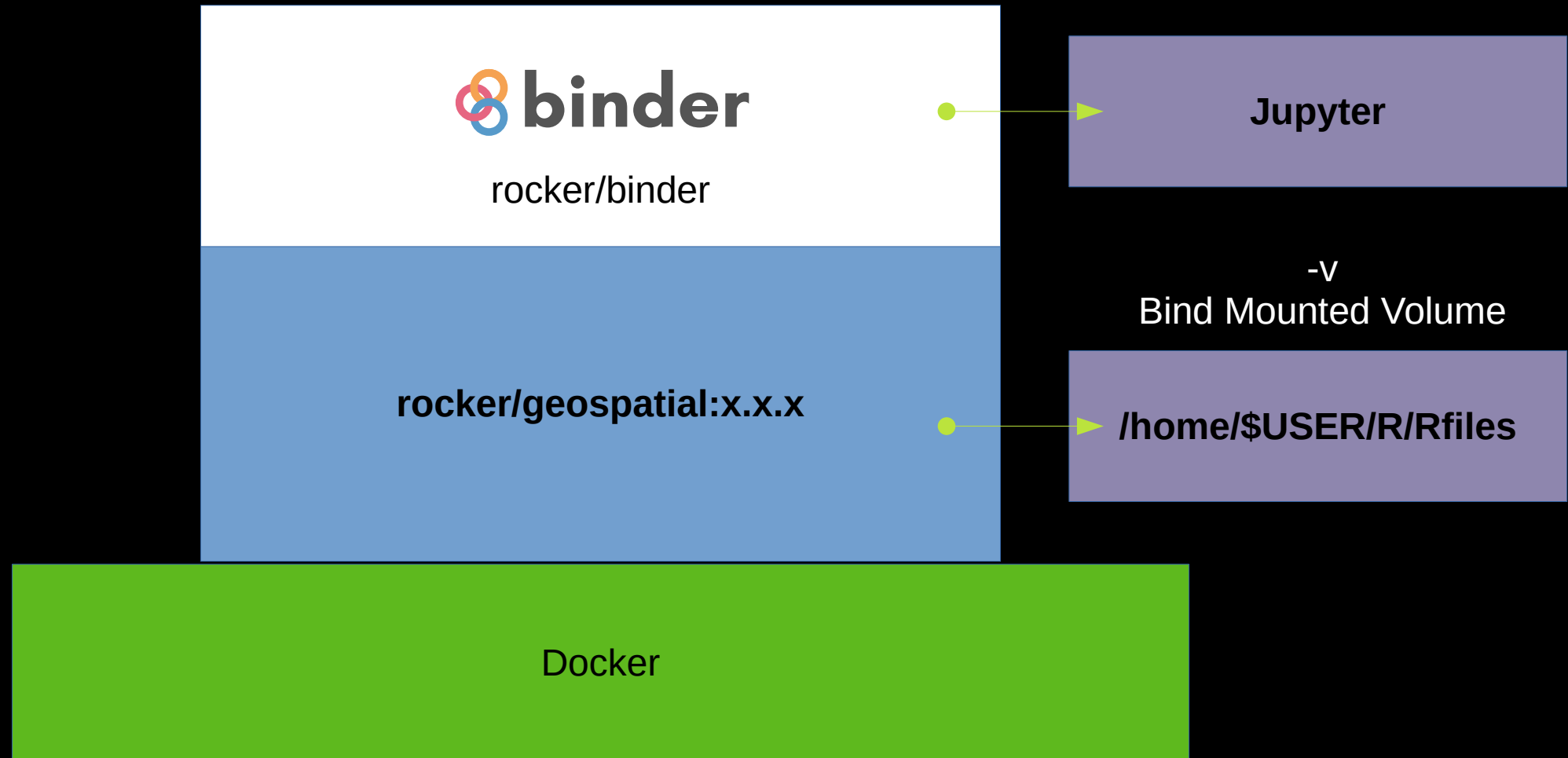
Docker

---

Building Up Our Docker Layers...



Building Up Our Docker Layers...



Building Up Our Docker Layers...



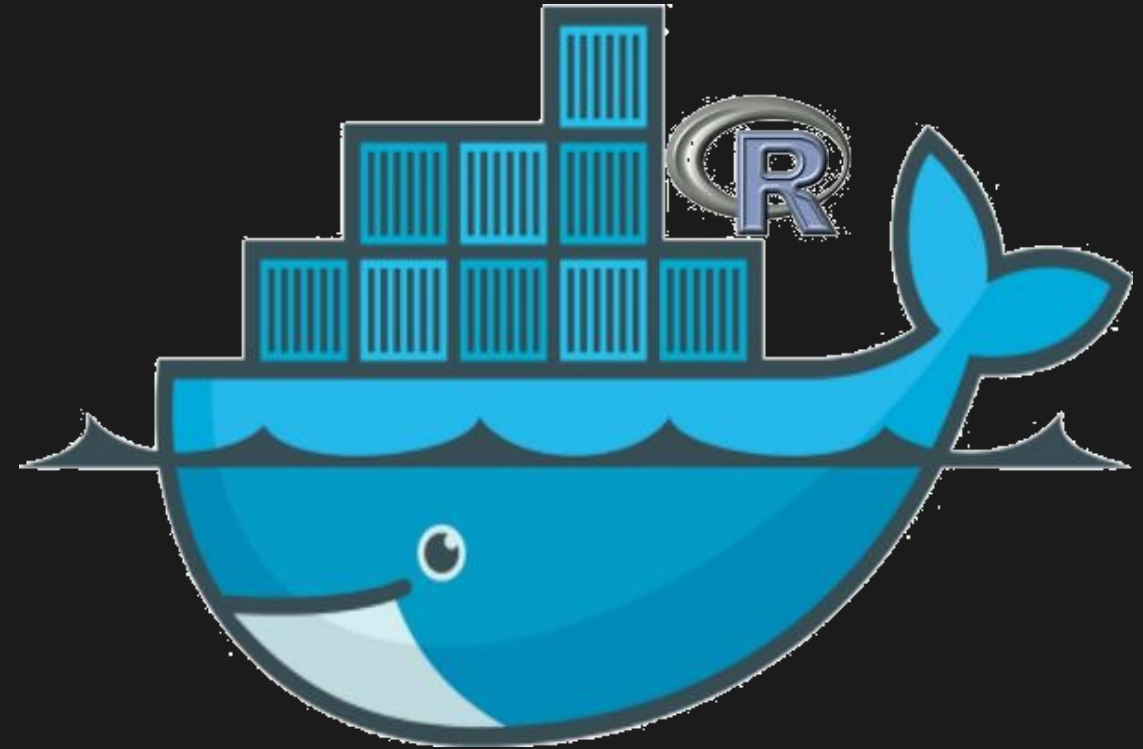
# Docker at 50,000 Feet

- Runs software packages called containers
- Developers can package up applications, including libraries and dependencies into a container



# Rocker at 50,000 Feet

- R on Docker = Rocker
- <https://www.rocker-project.org/images/>
- R with libraries and dependencies in nifty containers
- TidyVerse is a collection of packages that facilitates our visualizations



# Some Images in the Rocker Universe...

- <https://www.rocker-project.org/images/>

Image	Comments
r-ver	• Minimal
rstudio	• Adds rstudio. Great mid-sized container. Sans tidyverse
tidyverse	• Adds tidyverse. Great mid-sized container. Includes ggplot2 (data visualization stuff)
verse	• Adds TeX and publishing
geospatial	• Adds geospatial libraries. Large image

# Post-configuration Docker (Part 2 of 2)

- In RStudio:
  - `library(tidyverse)`
  - `library(devtools)`
  - `install.packages(ggfortify)` [or `install_github('sinhrks/ggfortify')`]
  - `install.packages("zoo")`
  - `library(ggfortify)`
  - `autoplot(AirPassengers)`

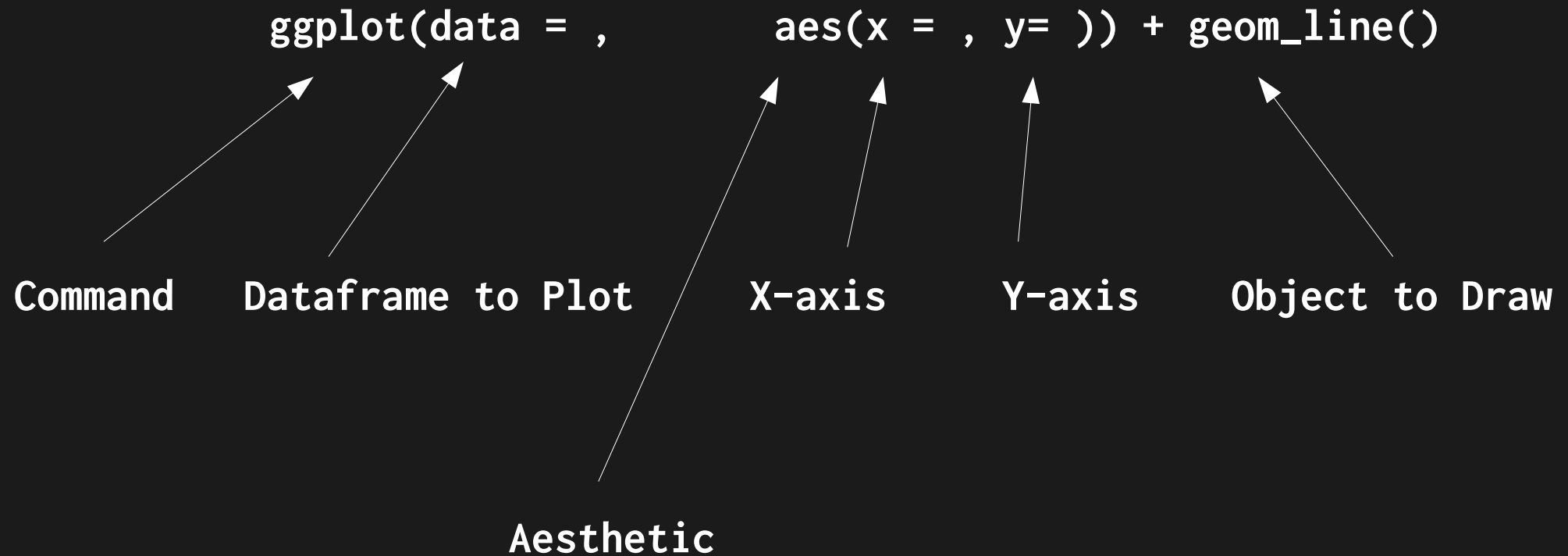
How Do I Save My Stuff?

---

# The Plot Thickens...

- We could use the basic plotting libraries, but use the ggplot2 libraries instead
- Why?
  - Flexible and versatile
    - Progressively modify through additional layers
  - Highly structured
    - Grammar of Graphics (beyond the scope of this talk)
  - Creates data visualizations
  - What datasets are available at our disposal?
    - `library(help = "datasets")`
    - `?(datasetname)`, e.g. `?mtcars`

# ggplot2 syntax



# Aesthetic Attributes

---

- Mapping data from dataframe to your plot's aesthetic attributes



# Example #1: Scatter Plot (x, y)

- `df <- trees`
- `trees`
  - List tree data
- `?trees`
  - Find out more info about this data
- `library("tidyverse")`
- `ggplot (df, aes (x=Girth, y=Volume)) + geom_point()`

## Example #2: Bar Chart (x)

- `mpg`
- `?mpg`
- `# geom_bar` is designed to make it easy to create bar charts that show
- `# counts (or sums of weights)`
- `g <- ggplot(mpg, aes(class))`
- `# Number of cars in each class:`
- `g + geom_bar()`

## Example #3: Bar Chart (x)

- `midwest`
- `?midwest`
- If there's no `y` specified, `ggplot` just counts the records
- `ggplot (data=midwest,aes(x=state))+geom_bar()`

# Example #4: Importing CSV Data and Line Chart

- `library("tidyverse")`
- `tsla_stock_metrics <- read_csv("https://www.sharpsightlabs.com/datasets/TSLA_start-to-2018-10-26_CLEAN.csv")`
- `print ("tsla_stock_metrics")`
- `ggplot(data=tsla_stock_metrics, aes(x=date,y=open_price))+geom_line()`

# Example #5: Histogram

- `ggplot(data = txhousing, aes(x = median)) + geom_histogram()`
- `geom_histogram` options:
  - `bins=100`
  - `color="purple"`
- `facet_wrap`:
  - `facet_wrap(~city)`

# Example #6: Density Plot

---

- `ggplot(data = txhousing, aes(x = median)) + geom_density()`

# What is this useful for?

---

- Use the histogram / density plots to:
  - Verify normal variable distribution
  - Detect outliers
  - Detect skewness
  - Look for interesting features in the data

What resources are available?

---



# Materials (Part 1 of 2)

Materials (detailed list)	Description	Link
R for Data Science Book	Hadley Wickham Reference	<a href="https://r4ds.had.co.nz/">https://r4ds.had.co.nz/</a>
Hands-On Programming with R	Garrett Grolmund R-specific Learn By Doing Book	<a href="http://shop.oreilly.com/product/0636920028574.do">http://shop.oreilly.com/product/0636920028574.do</a>
Various R related cheatsheets	R cheatsheets	<a href="https://www.rstudio.com/resources/cheatsheets/">https://www.rstudio.com/resources/cheatsheets/</a>

# Materials (Part 1 of 2)

Materials (detailed list)	Description	Link
Datacamp Subscription	Datacamp Training Courses	<a href="https://www.datacamp.com/home">https://www.datacamp.com/home</a>
ggplot2 Book	Hadley Wickham ggplot Reference	<a href="https://www.amazon.com/dp/0387981403/ref=cm_sw_su_dp?tag=ggplot2-20">https://www.amazon.com/dp/0387981403/ref=cm_sw_su_dp?tag=ggplot2-20</a>

# Websites (1 of 1)

Materials (detailed list)	Description	Link
Sharpsight Tutorial	Beginner Tips from Sharpsight Labs	<a href="https://www.sharpsightlabs.com/blog/ggplot2-tutorial/">https://www.sharpsightlabs.com/blog/ggplot2-tutorial/</a>
Sharpsight Tutorial	Histogram and Density Plot	<a href="https://www.sharpsightlabs.com/blog/ggplot-histogram/">https://www.sharpsightlabs.com/blog/ggplot-histogram/</a>
R-Statistics Tutorial	Complete ggplot2 Tutorial (3 parts!)	<a href="http://r-statistics.co/Complete-Ggplot2-Tutorial-Part1-With-R-Code.html">http://r-statistics.co/Complete-Ggplot2-Tutorial-Part1-With-R-Code.html</a>

Questions???

---

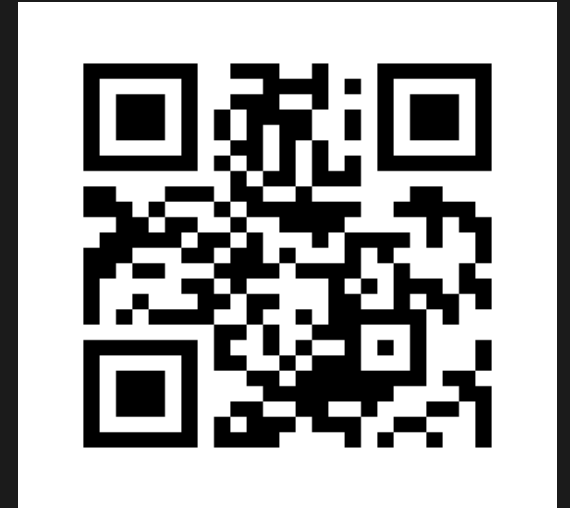
# Recap

---

- Learned some charting and plotting techniques

# Perry Rivera

- E-mail: [perry-rivera@alumni.calpoly.edu](mailto:perry-rivera@alumni.calpoly.edu)
- LinkedIn: [@pvrconsulting](#)
- Slides:
  - Part 1: <https://tinyurl.com/yxe3q23y>
  - Part 2: <https://tinyurl.com/y5os9wl2>
- Github:
  - [https://github.com/perryrivera/r\\_development\\_presentation](https://github.com/perryrivera/r_development_presentation)





# Thank You!