

VIVIAN PENG

Data Visualization Workshop

R Conference 2020



Agenda

Icebreaker & Intros

Overview of data viz

Objectives

Intro to Shiny

Intros

Name

Where are
you based?

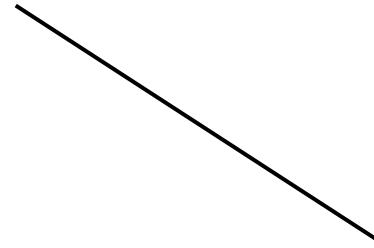
What do you do
for work?

Goals for this
workshop

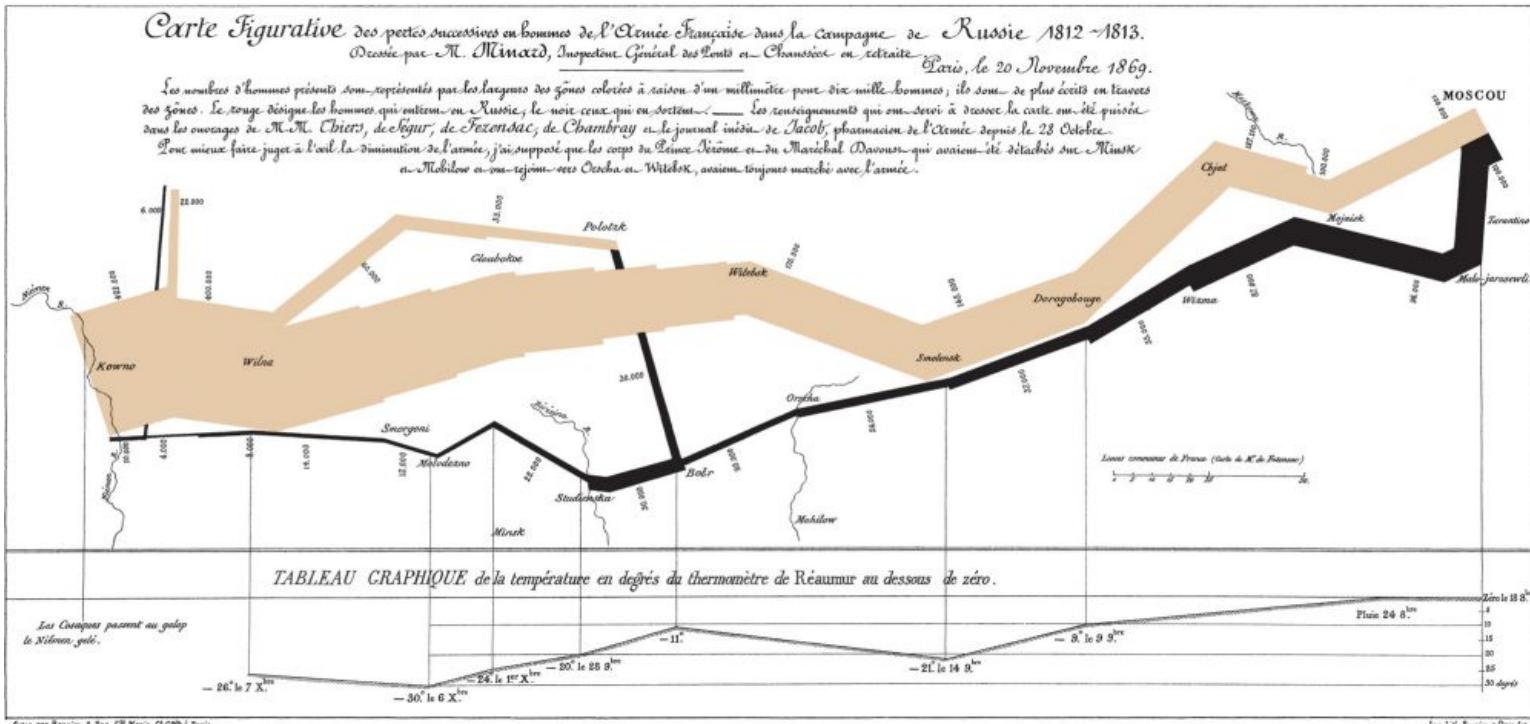
What's your
hidden talent?

exploratory

explanatory



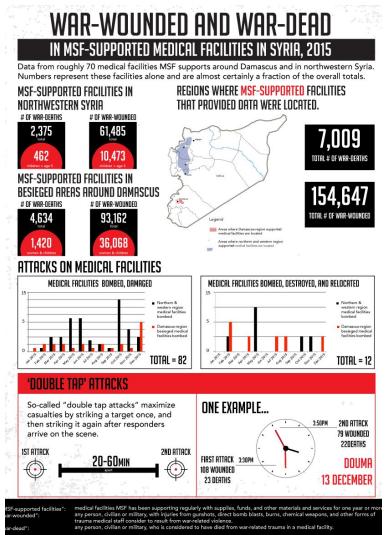
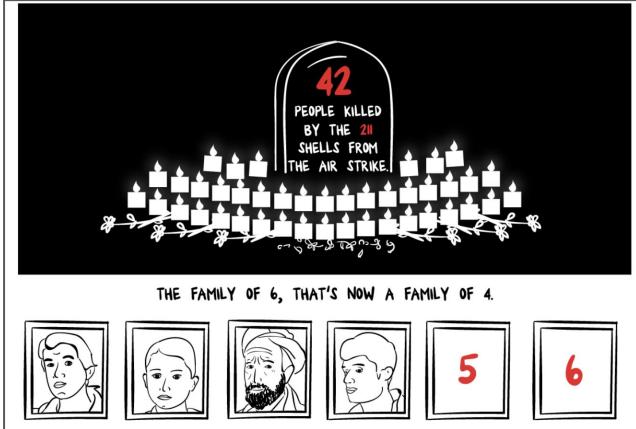
Data Visualizations



visual impact

data rich





B U I L D I N G

B L O C K S

Objectives

Build a Shiny app

Load your own data

Understand Reactivity

Deploy to shinyapps.io

Design-Thinking

Shiny

Why Shiny?

Interactive is helpful!

Reproducible

Web-based, don't have to load anything

Good for exploratory analysis

1

Try code in
regular R file first

2

Then add to a
Shiny app



Simple Shiny App Setup

app.R

```
library(shiny)  
  
ui <- fluidPage(  
  )
```

Tells Shiny what to display on the webpage

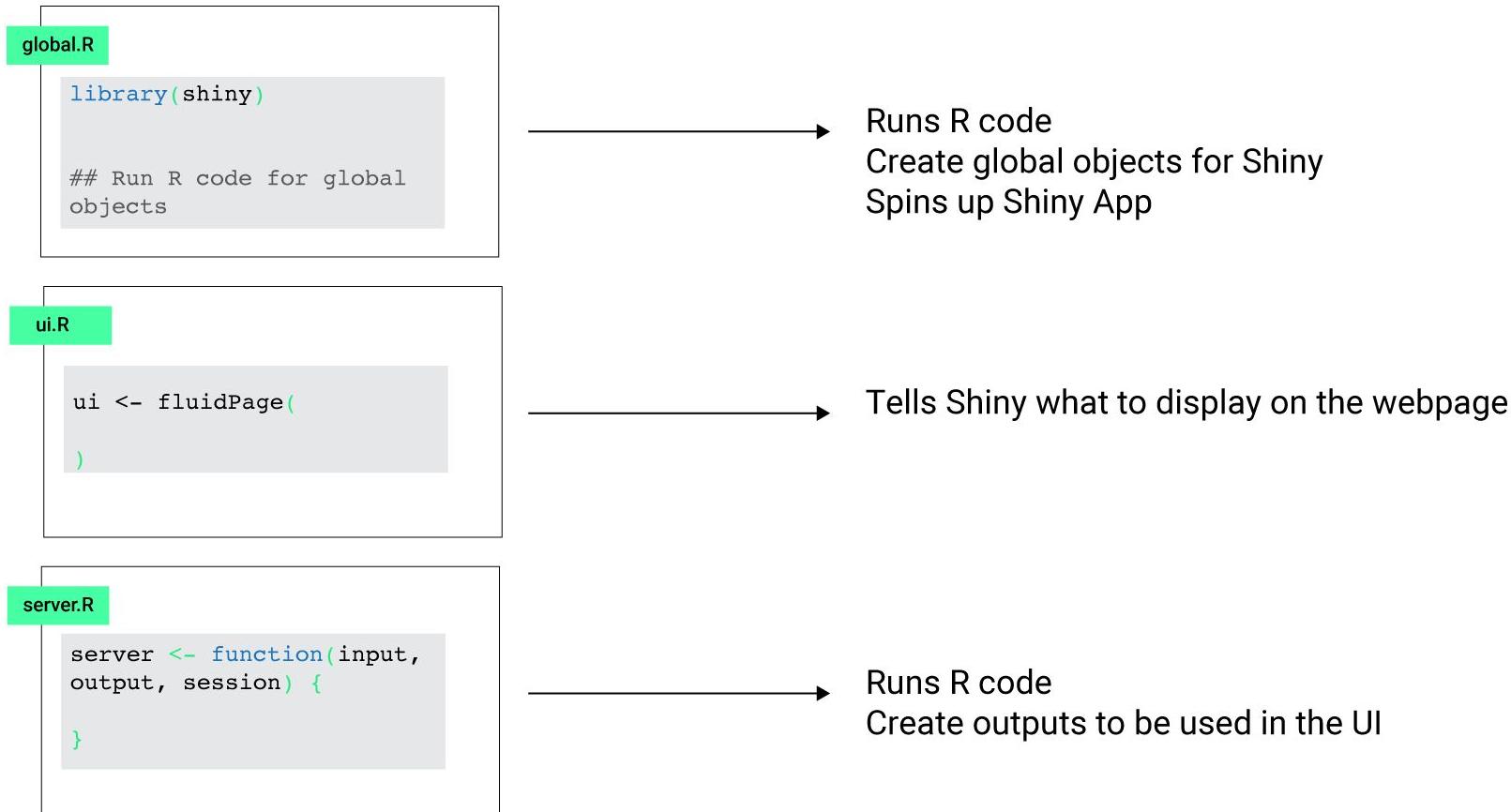
```
server <- function(input,  
output, session) {  
  }
```

Runs R code
Create outputs to be used in the UI

```
shinyApp(ui, server)
```

Spins up Shiny App

Advanced Shiny App Setup



Simple Shiny App Setup

app.R

```
library(shiny)

ui <- fluidPage(
)
```

Tells Shiny what to display on the webpage

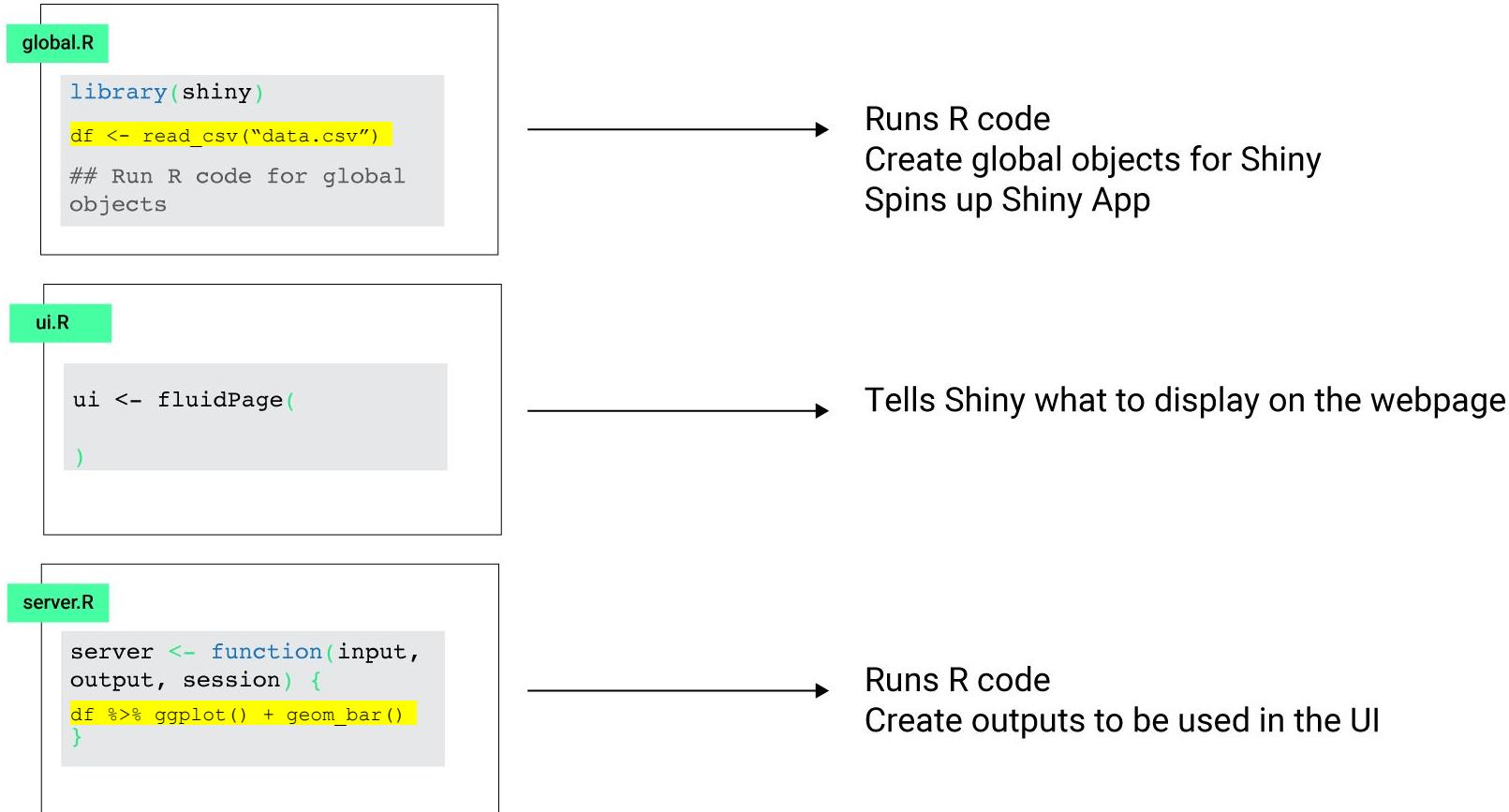
```
server <- function(input,
output, session) {
df <- read_csv("data.csv")
df %>% ggplot() + geom_bar()
```

Runs R code
Create outputs to be used in the UI

```
shinyApp(ui, server)
```

Spins up Shiny App

Advanced Shiny App Setup



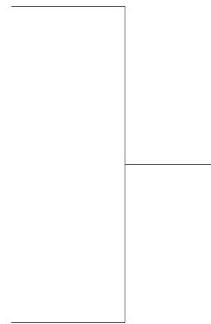
Let's try it!

1. Open R project
2. Create a new R file

3. | + Shift + Tab

ui section

```
ui <- fluidPage(  
  h1(),  
  fluidRow(),  
  sliderInput(),  
  textOutput()  
)
```



various types of elements

each separated by a comma

no comma on the last one!

[See all layout options here](#)

[See all HTML tags](#)

[See all Shiny widgets](#)

server section

```
server <- function(input, output, session){  
  
  df <- read_csv("data.csv")  
  
  output$math <- renderText({2+2})  
  
}
```

regular R code

NOT separated by commas!

ui element

DTOutput
imageOutput
plotOutput
tableOutput
textOutput
uiOutput

+

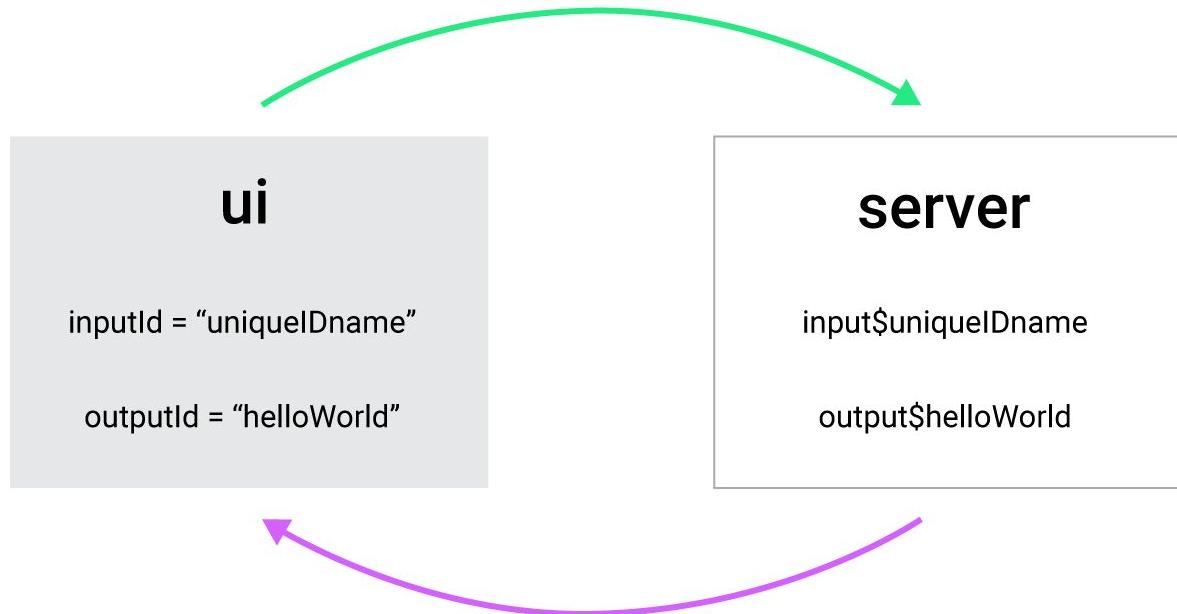
server function

renderDT
renderImage
renderPlot
renderTable
renderText
renderUI

=

creates

DataTable
Raw HTML
Image
Table
Text
Raw HTML



ui section

```
selectInput(  
  inputID = "barX",  
  label = "Pick an animal",  
  choices=c("cat", "bear", "dog")  
,  
  
textOutput(outputId = "math")
```

server section

```
output$plot <- renderPlot({  
  ggplot(df,aes(x=input$barX,  
  y = total)  
})  
  
output$math <- renderText({ 2+2 })
```

Add UI element to your interface

Tell shiny how to build the object

ui section

```
selectInput(  
  inputID = "barX",  
  label = "Pick an animal",  
  choices=c("cat", "bear", "dog")  
,  
  
textOutput(outputId = "math")
```

server section

```
output$plot <- renderPlot({  
  ggplot(df,aes(x=input$barX,  
  y = total)  
})  
  
output$math <- renderText({ 2+2 })
```

Add UI element to your interface

Tell shiny how to build the object

open 01_intro/app.R



provided by **lyft**

`open 02_project/eda.R`

open 02_project/app.R

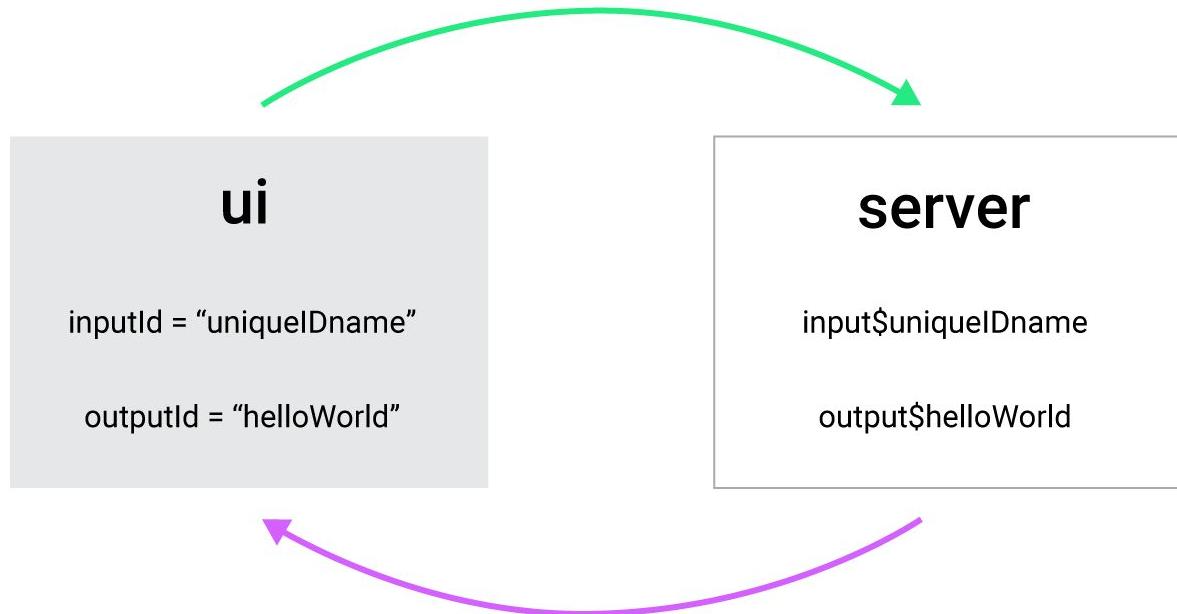
Why interactive?

Helpful for exploratory analysis

Gives user control

Allows for you to design with audience in mind

Dynamic



Make interactive plots

1. Add UI element with input choices
2. Update plot code with new input variable

```
ggplot(df, aes(y=col_name))
```

static

```
ggplot(df, aes(y=input$var))
```

interactive

1

```
ui <- fluidPage(  
  
  selectInput(inputId="bar_yaxis",  
              label = h3("Select Variable"),  
              choices=c("Avg Duration" = "avg_duration", "Avg Age" = "avg_age", "Total  
                        Rides"= "total_rides")),  
  
  plotOutput("bar_plot")  
  
)  
  
server <- function(input, output, session) {  
  
  output$bar_plot <- renderPlot({  
    ggplot(summary,aes(y=summary[[input$bar_yaxis]], x=start_station_name))  
  })  
  
}
```

2

Ways to subset a dataframe

Base R

```
df$column
```

```
df[ [column] ]
```

Interactive Shiny

```
df$input$column
```

```
df[ [input$column] ]
```

This will break in
Shiny because of
how we use "\$"
to reference
input/output ID's

open 02_project/app.R

Exercise

Create reactive y label

plotly

Deploy to shinyapps.io!

Deploy (Part 1)

*Only have to do this once!

1. Sign up for [shinyapps.io](#)
2. `install.packages('rsconnect')`
3. Close RStudio
4. Re-open RStudio
5. Run this chunk of code in console

STEP 2 - AUTHORIZE ACCOUNT

The `rsconnect` package must be authorized to your account using a token and secret. To do this, click the copy button below and we'll copy the whole command you need to your clipboard. Just paste it into your console to authorize your account. Once you've entered the command successfully in R, that computer is now authorized to deploy applications to your shinyapps.io account.

```
rsconnect::setAccountInfo(name='your_username_here',
                           token='ABC812390123ASDFCSD098139405DFKLA',
                           secret='<SECRET>')
```

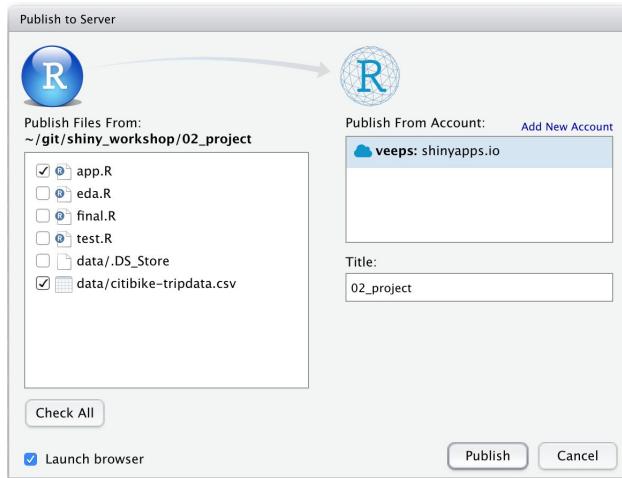
[Show secret](#)

[Copy to clipboard](#)

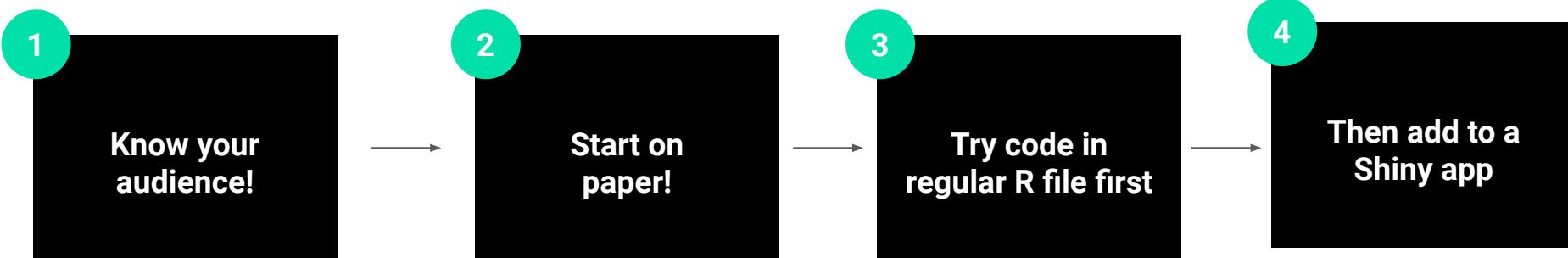
In the future, you can manage your tokens from the [Tokens](#) page the settings menu.

Deploy (Part 2)

1. Open your app.R file
2. Run App
3. Click on “Publish” button on top right
4. Only check files needed



Let's take a design break



In data science, we spend the majority of our time **cleaning data.**

In the design world, we spend majority of our time **understanding the problem & thinking through concepts.**

But it's not a linear process...

Johnson & Johnson

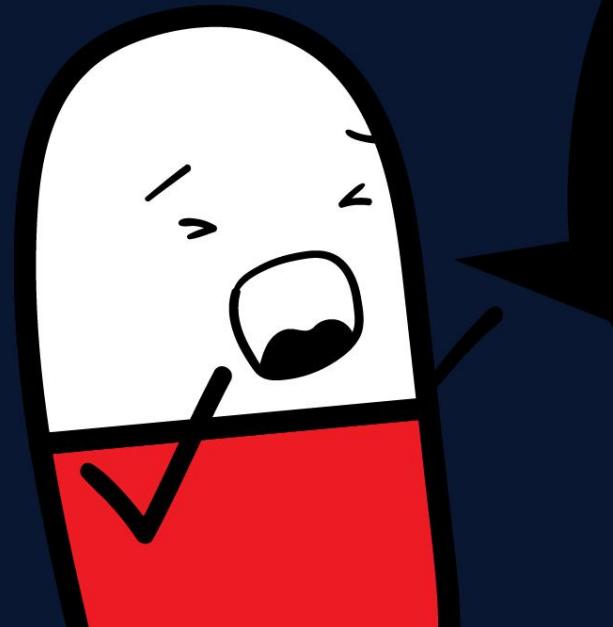
CAN YOU HEAR US NOW?



ACCESS
CAMPAIGN
 MÉDECINS
SANS FRONTIÈRES



Johnson & Johnson

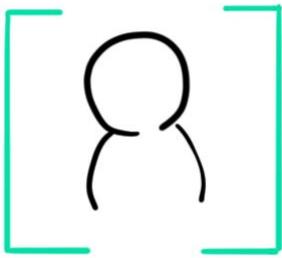


CAN YOU
HEAR US?



**But there is a creative process,
and we have ways to document it!**

3 Questions to Help You Understand Your Audience



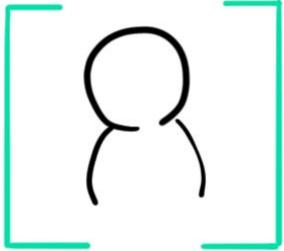
Who is your primary audience?



How do you want them to feel?



What action do they take?



Primary Audience

- What do they care about?**
- What problem are they trying to solve?**
- What limitations are they working with?**
(time, tech barriers, data literacy, bureaucracy, influence)



All the Feels

- Understand what kinds of emotions are they walking into this project with?

- When they walk away from this, how do you want them to feel?
(Ex: At ease, or angered to take action, or happy with the opportunity)

If you're not sure, choose happy.





Take Action

- Is this a one time action v. ongoing?
- What can they control v. who do they need to influence?
- What's the takeaway message?

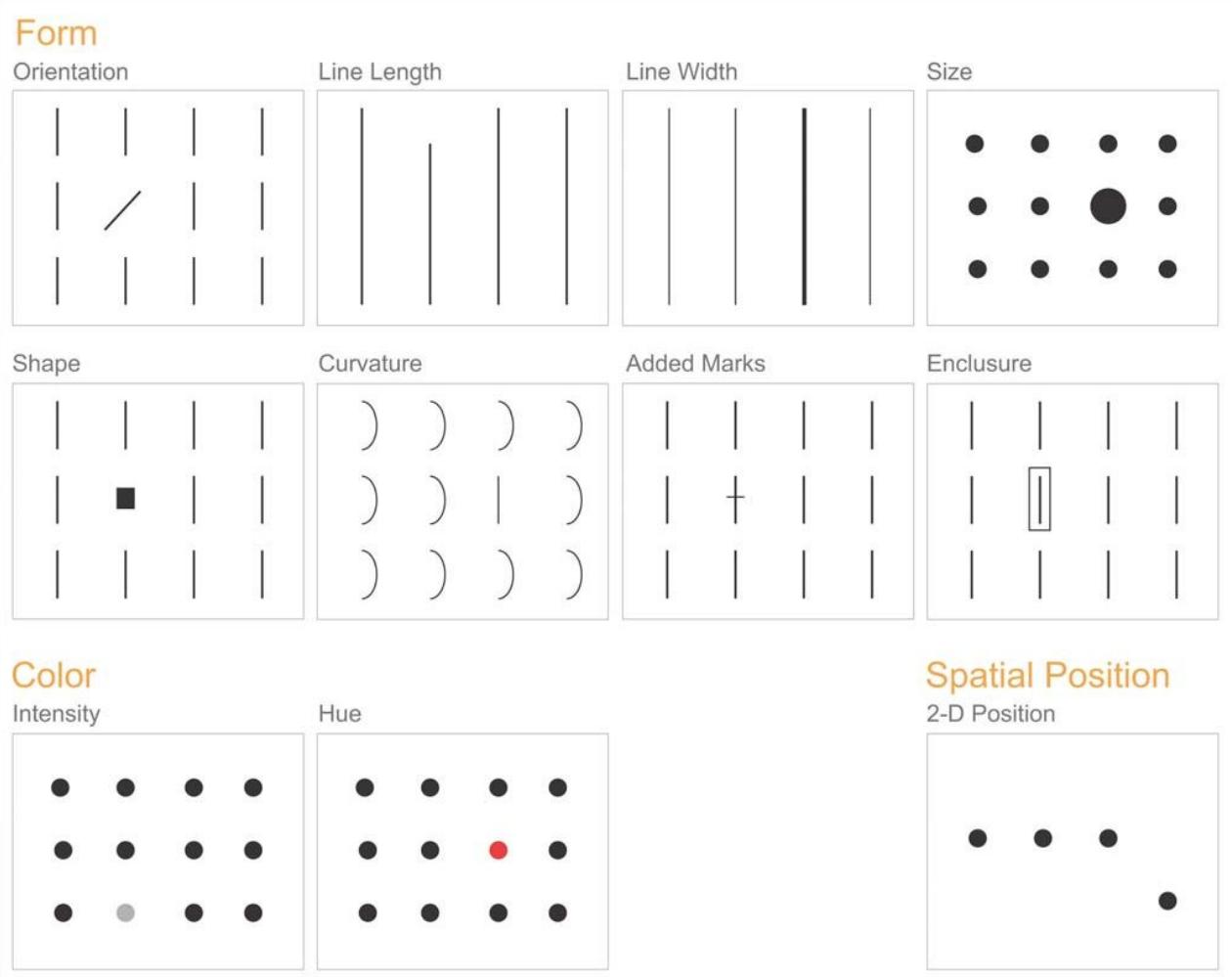
Audience

Who is our primary audience?	How do we want them to feel?	What action do they take?
Execs (limited time, uncomfortable with data, trying to solve ____)	At ease with the data and analysis	Comfortable with our recommendation (taking a risk)
Data Folks (deeply invested, highly critical, trying to solve ____)	Confident about the data collection & analysis process	Create recommendations Communicate to stakeholders
Public Audience (limited time, may not care about this)	Angered Empathetic	Share this information Sign up to learn more
Marketing Folks ()		

Visual Cues

* From [Storytelling with Data](#) by Cole Nussbaumer Knaflic

Only use 1-2 cues per graphic!



LOS ANGELES COUNTY KEY METRICS

Key Takeaways:

The following metrics have seen **undesirable** movement over the past seven days for which data are available:

New Cases (vs baseline), **New Hospitalizations** (vs baseline), **Acute Care Bed Capacity**, **ICU Capacity**, and **Unemployment Insurance** (vs baseline)

The following metrics have seen **desirable** movement over the past seven days: **New Cases** (vs last week), **New Deaths**, **COVID Positive Test Rate** (vs baseline), **New Hospitalizations** (vs last week), **Unemployment Insurance** (vs last week), and **Crime**

- ➡ Undesired Trend
- ➡ Desired Trend
- Desired Change
- Undesired Change

Source: Los Angeles County Department of Health, COVID Tracking Project

Prepared by Mayor's Innovation Team

Page 6

[source](#)

METRICS As of Aug 7, 2020, 1 PM	TODAY'S AVERAGES	CURRENT TREND VS LAST WEEK	% CHANGE VS LAST WEEK	CURRENT TREND VS BASELINE	% CHANGE VS BASELINE
New Cases (7-day avg.)	2,295 (7-day avg)	➡	-20.4%	➡	+120.1%
New Deaths (7-day avg.)	44 (7-day avg)	➡	-6.3%	➡	-12.1%
COVID Cumulative Positive Test Rate	9.9% (overall)	↔	NO CHANGE COVID+ Test Rate is % Difference w/ last week	➡	-4.8% COVID+ Test Rate is % Difference w/ baseline
New Hospitalizations (7-day avg)	156 (7-day avg)	➡	-8.0%	➡	+0.1%
Acute Care Bed Capacity (7-day avg)	499 (7-day avg)	➡	-11.6%	➡	-55.3%
ICU Bed Capacity (7-day avg)	121 (7-day avg)	➡	-12.7%	➡	-50.3%
LA City Crime (7-day avg)* baseline = 2019 YTD	463 daily crimes (7-day avg)	➡	-4.7%	➡	-26.5%
LA City Unemployment Ins. Claims baseline = equivalent 2019 week	33,988 (as of 7/25/2020)	➡	-5.5%	➡	+1530%

LOS ANGELES COUNTY KEY METRICS

Key Takeaways:

The following metrics have seen **undesirable** movement over the past seven days for which data are available:

New Cases (vs baseline), **New Hospitalizations** (vs baseline), **Acute Care Bed Capacity**, **ICU Capacity**, and **Unemployment Insurance** (vs baseline)

The following metrics have seen **desirable** movement over the past seven days: **New Cases** (vs last week), **New Deaths**, **COVID Positive Test Rate** (vs baseline), **New Hospitalizations** (vs last week), **Acute Care Bed Capacity** (vs baseline), **ICU Bed Capacity** (vs baseline), **LA City Crime** (vs baseline), and **LA City Unemployment Insurance Claims** (vs baseline).

3 Cues Used Here:

- Line slope (orientation)
- Line color (hue)
- Background color (hue)

→ Undesired Trend

→ Desired Trend

Desired Change

Undesired Change

Source: Los Angeles County Department of Health, COVID Tracking Project

Prepared by Mayor's Innovation Team

Page 6

[source](#)

METRICS As of Aug 7, 2020, 1 PM	TODAY'S AVERAGES	CURRENT TREND VS LAST WEEK	% CHANGE VS LAST WEEK	CURRENT TREND VS BASELINE	% CHANGE VS BASELINE
New Cases (7-day avg.)	2,295 (7-day avg)	↓	-20.4%	↑	+120.1%
New Deaths (7-day avg.)	44 (7-day avg)	↓	-6.3%	↓	-12.1%
COVID Cumulative Positive Test Rate	9.9% (overall)	↔	NO CHANGE COVID+ Test Rate is % Difference w/ last week	↓	-4.8% COVID+ Test Rate is % Difference w/ baseline
New Hospitalizations (7-day avg)	156 (7-day avg)	↓	-8.0%	↑	+0.1%
Acute Care Bed Capacity (7-day avg)	499 (7-day avg)	↓	-11.6%	↓	-55.3%
ICU Bed Capacity (7-day avg)	121 (7-day avg)	↓	-12.7%	↓	-50.3%
LA City Crime (7-day avg)* baseline = 2019 YTD	463 daily crimes (7-day avg)	↓	-4.7%	↓	-26.5%
LA City Unemployment Ins. Claims baseline = equivalent 2019 week	33,988 (as of 7/25/2020)	↓	-5.5%	↑	+1530%

LOS ANGELES COUNTY KEY METRICS

Key Takeaways:

The following metrics have seen **undesirable** movement over the past seven days for which data are available:

New Cases (vs baseline), **New Hospitalizations** (vs baseline), **Acute Care Bed Capacity**, **ICU Capacity**, and **Unemployment Insurance** (vs baseline)

The following metrics have seen **desirable** movement over the past seven days: **New Cases** (vs last week), **New Deaths**, **COVID Positive Test Rate** (vs baseline), **New Hospitalizations** (vs last week), **Unemployment Insurance** (vs last week), and **Crime**

- Undesired Trend
- Desired Trend
- Desired Change
- Undesired Change

Source: Los Angeles County Department of Health, COVID Tracking Project

Prepared by Mayor's Innovation Team

Page 6

[source](#)

METRICS As of Aug 7, 2020, 1 PM	TODAY'S AVERAGES	CURRENT TREND VS LAST WEEK	% CHANGE VS LAST WEEK	CURRENT TREND VS BASELINE	% CHANGE VS BASELINE
New Cases (7-day avg.)	2,295 (7-day avg)	↓	-20.4%	↑	+120.1%
New Deaths (7-day avg.)	44 (7-day avg)	↓	-6.3%	↓	-12.1%
COVID Cumulative Positive Test Rate	9.9% (overall)	↔	NO CHANGE COVID+ Test Rate is % Difference w/ last week	↓	-4.8% COVID+ Test Rate is % Difference w/ baseline
New Hospitalizations (7-day avg)	156 (7-day avg)	↓	-8.0%	↑	+0.1%
Acute Care Bed Capacity (7-day avg)	499 (7-day avg)	↓	-11.6%	↓	-55.3%
ICU Bed Capacity (7-day avg)	121 (7-day avg)	↓	-12.7%	↓	-50.3%
LA City Crime (7-day avg)* baseline = 2019 YTD	463 daily crimes (7-day avg)	↓	-4.7%	↓	-26.5%
LA City Unemployment Ins. Claims baseline = equivalent 2019 week	33,988 (as of 7/25/2020)	↓	-5.5%	↑	+1530%

LOS ANGELES COUNTY KEY METRICS

Key Takeaways:

The following metrics have seen **undesirable** movement over the past seven days for which data are available:

New Cases (vs baseline), **New Hospitalizations** (vs baseline), **Acute Care Bed Capacity**, **ICU Capacity**, and **Unemployment Insurance** (vs baseline)

The following metrics have seen **desirable** movement over the past seven days: **New Cases** (vs last week), **New Deaths**, **COVID Positive Test Rate** (vs baseline), **New Hospitalizations** (vs last week), **Unemployment Insurance** (vs last week), and **Crime**

- ➡ Undesired Trend
- ➡ Desired Trend
- Desired Change
- Undesired Change

Source: Los Angeles County Department of Health, COVID Tracking Project

Prepared by Mayor's Innovation Team

Page 6

[source](#)

METRICS As of Aug 7, 2020, 1 PM	TODAY'S AVERAGES	CURRENT TREND VS LAST WEEK	% CHANGE VS LAST WEEK	CURRENT TREND VS BASELINE	% CHANGE VS BASELINE
New Cases (7-day avg.)	2,295 (7-day avg)	➡	-20.4%	➡	+120.1%
New Deaths (7-day avg.)	44 (7-day avg)	➡	-6.3%	➡	-12.1%
COVID Cumulative Positive Test Rate	9.9% (overall)	↔	NO CHANGE COVID+ Test Rate is % Difference w/ last week	➡	-4.8% COVID+ Test Rate is % Difference w/ baseline
New Hospitalizations (7-day avg)	156 (7-day avg)	➡	-8.0%	➡	+0.1%
Acute Care Bed Capacity (7-day avg)	499 (7-day avg)	➡	-11.6%	➡	-55.3%
ICU Bed Capacity (7-day avg)	121 (7-day avg)	➡	-12.7%	➡	-50.3%
LA City Crime (7-day avg)* baseline = 2019 YTD	463 daily crimes (7-day avg)	➡	-4.7%	➡	-26.5%
LA City Unemployment Ins. Claims baseline = equivalent 2019 week 7/25/2020	33,988 (as of 7/25/2020)	➡	-5.5%	➡	+1530%

LOS ANGELES COUNTY KEY METRICS

Key Takeaways:

The following metrics have seen **undesirable** movement over the past seven days for which data are available: **New Cases** (vs baseline), **New Deaths** (vs last week), **COVID+ Positive Test Rate** (vs last week), **New Hospitalizations** (vs baseline), **Acute Care Bed Capacity** (vs baseline), **ICU Capacity** (vs baseline), and **Unemployment Insurance** (vs baseline)

The following metrics have seen **desirable** movement over the past seven days: **New Cases** (vs last week), **New Deaths** (vs last week), **COVID+ Positive Test Rate** (vs baseline), **New Hospitalizations** (vs last week), **Acute Care Bed Capacity** (vs last week), **Unemployment Insurance** (vs last week), and **Crime**



Source: Los Angeles County Department of Health, COVID Tracking Project

Prepared by Mayor's Innovation Team

Page 1

METRICS As of July 29, 2020, 1 PM	TODAY'S NUMBERS	% CHANGE VS LAST WEEK	% CHANGE VS BASELINE
New Cases (7-day avg.)	2,695 (7-day avg)	-15.3%	+158.4%
New Deaths (7-day avg.)	45 (7-day avg)	+7.9%	-11.0%
COVID Cumulative Positive Test Rate	10.1% (overall)	+0.4% <small>COVID+ Test Rate is % Difference w/ last week</small>	-4.6% <small>COVID+ Test Rate is % Difference w/ baseline</small>
New Hospitalizations (7-day avg)	169 (7-day avg)	-11.9%	+8.4%
Acute Care Bed Capacity (7-day avg)	549 (7-day avg)	+6.7%	-50.8%
ICU Bed Capacity (7-day avg)	130 (7-day avg)	+30.0%	-46.4%
LA City Crime (7-day avg)* baseline = 2019 YTD	477 daily crimes (7-day avg)	-6.7%	-22.7%
LA City Unemployment Ins. Claims baseline = equivalent 2019 week	27,809 (as of 7/4/2020)	-1.1%	+1055%

- Don't walk into something new and start changing everything. Understand how it's being used first, and who uses this information.
- Ways to engage stakeholders in change process:

Instead of saying...

Try saying...

Why?

Can you help me understand how this is being used?

I think it's better to...

How about if we tried.... Would it make it difference if we

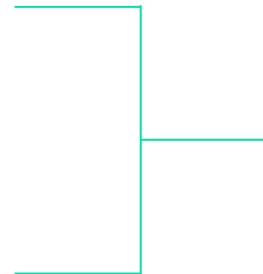
I don't like...

Tell me what you like about this? What's working?

Color is the easiest cue to change.
How do might we use it with *intention*?

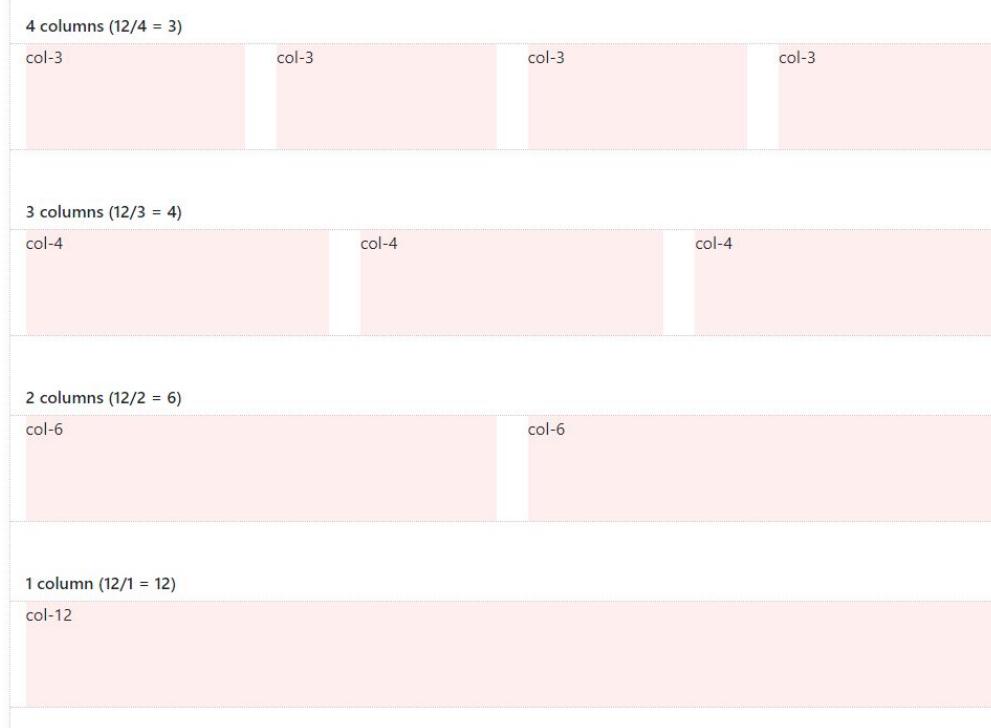
Some UI frameworks

Output function	Creates
<code>fluidPage()</code>	Responsive design
<code>fixedPage()</code>	Fixed page
<code>fillPage()</code>	Takes up full width
<code>navbarPage()</code>	Navigation bar at top. Uses tabPanel() and navbarMenu()



Try one of these
frameworks and
see how it
changes your
layout!

fluidPage() Uses Bootstrap Grid



`fluidRow()`

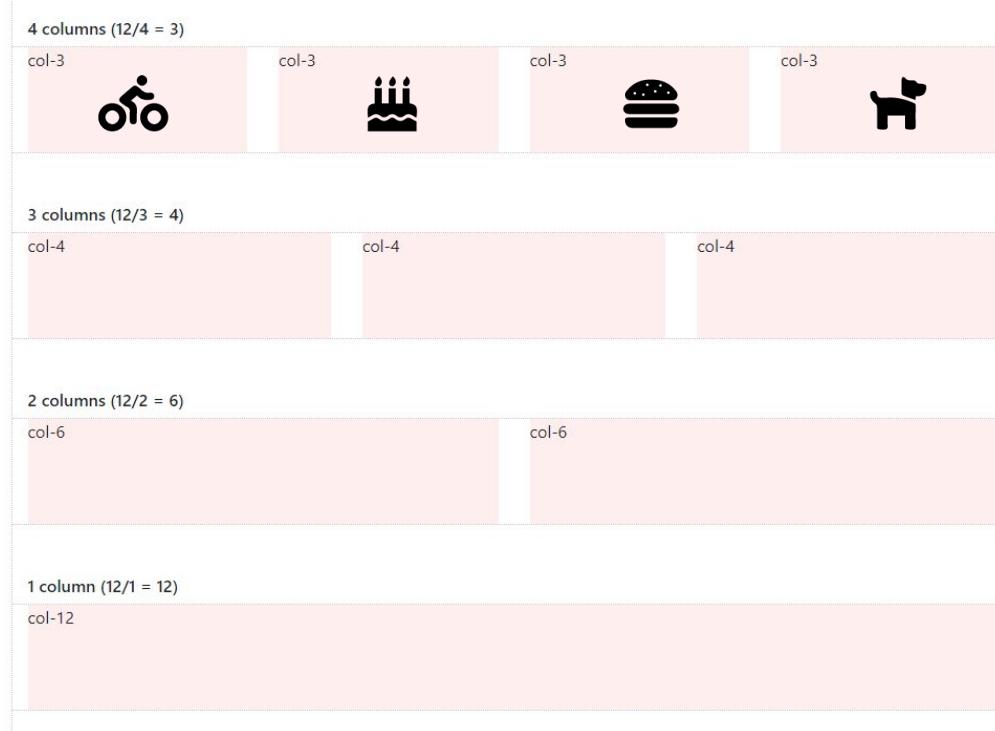
`fluidRow()`

`fluidRow()`

`fluidRow()`

[Image source](#)

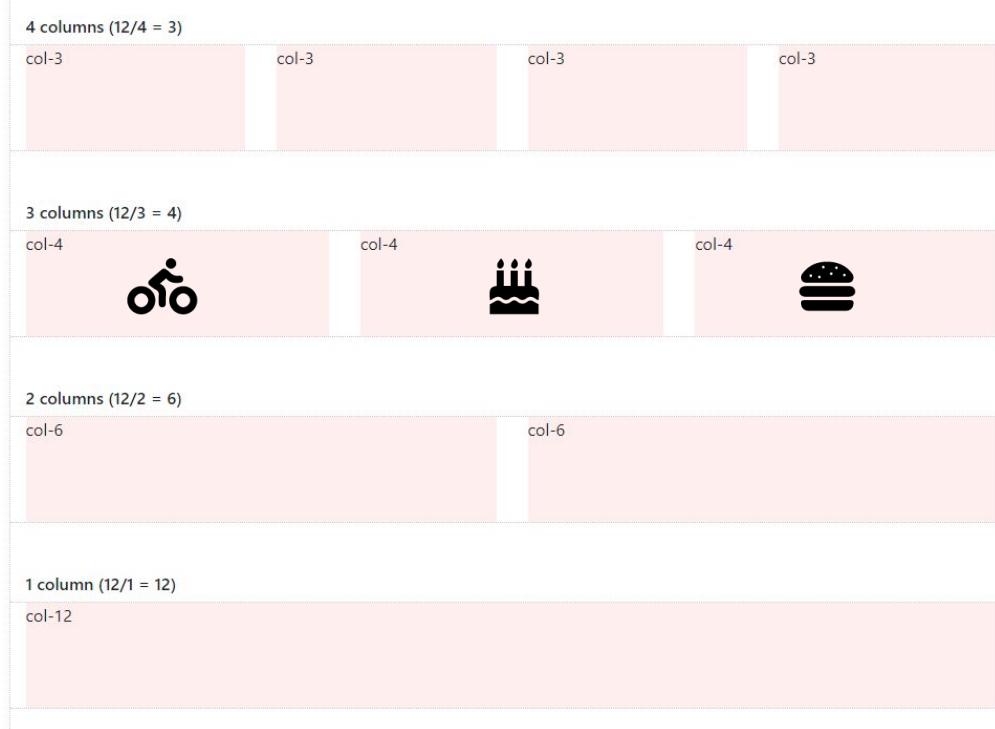
fluidPage() Uses Bootstrap Grid



```
fluidRow(  
  column(3, ),  
  column(3, ),  
  column(3, ),  
  column(3, ))
```

Image source

fluidPage() Uses Bootstrap Grid



```
fluidRow(  
  column(4, ) ,  
  column(4, ) ,  
  column(4, )  
)
```

[Image source](#)

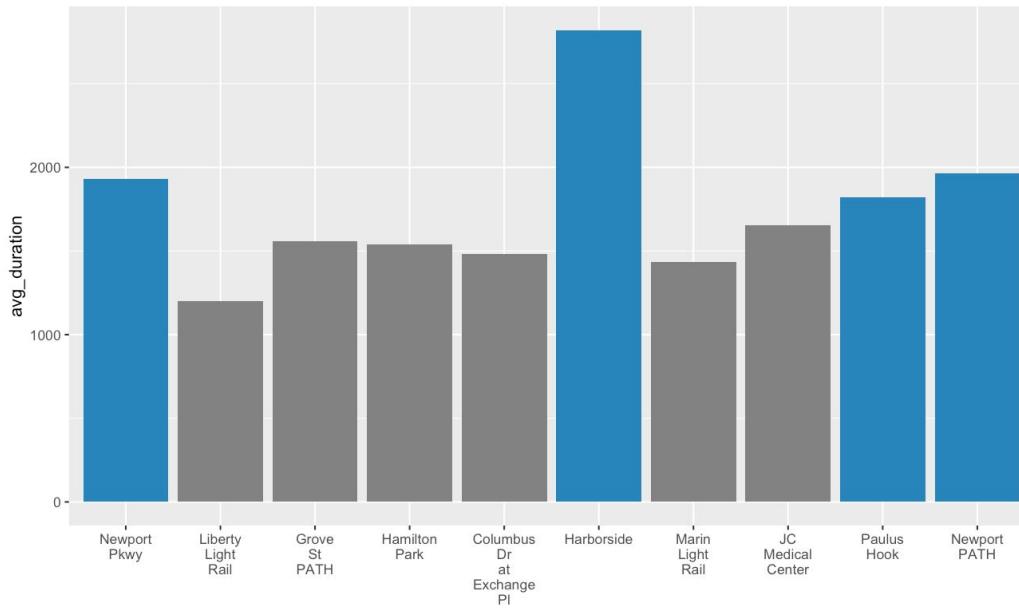
How do we create this?

col-3

col-9

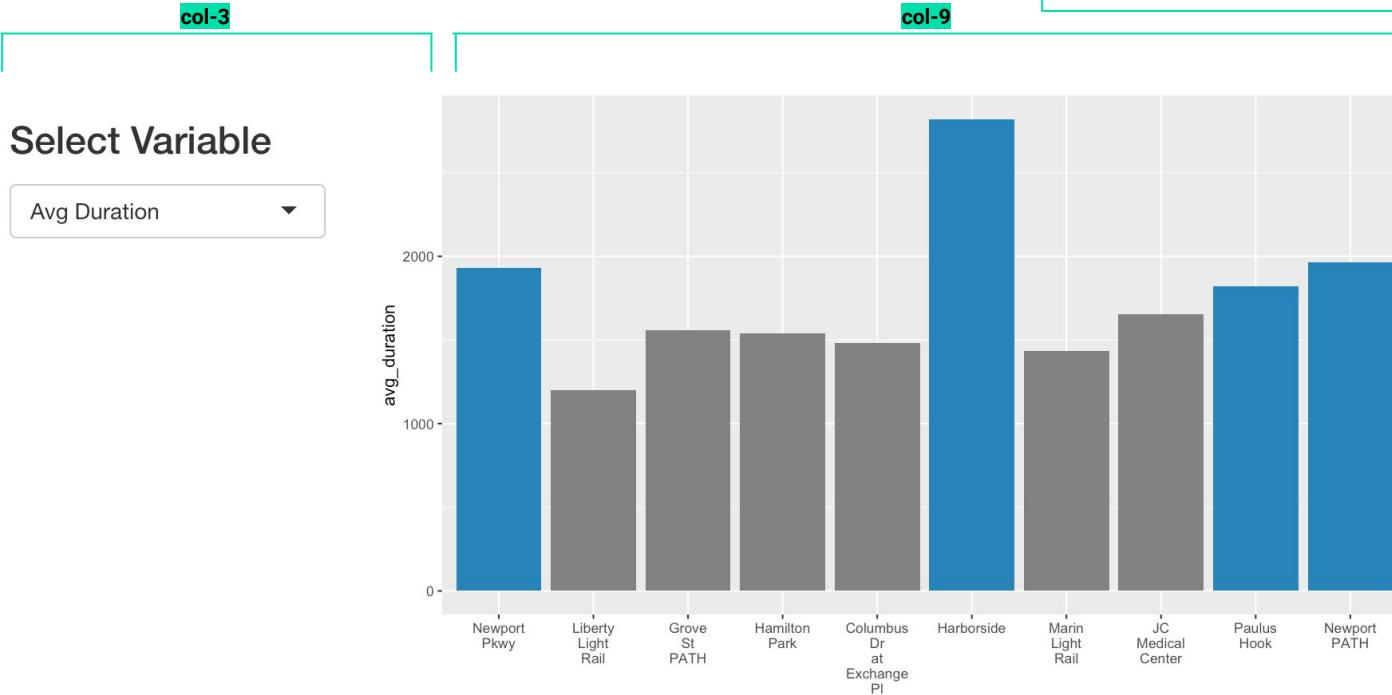
Select Variable

Avg Duration ▾



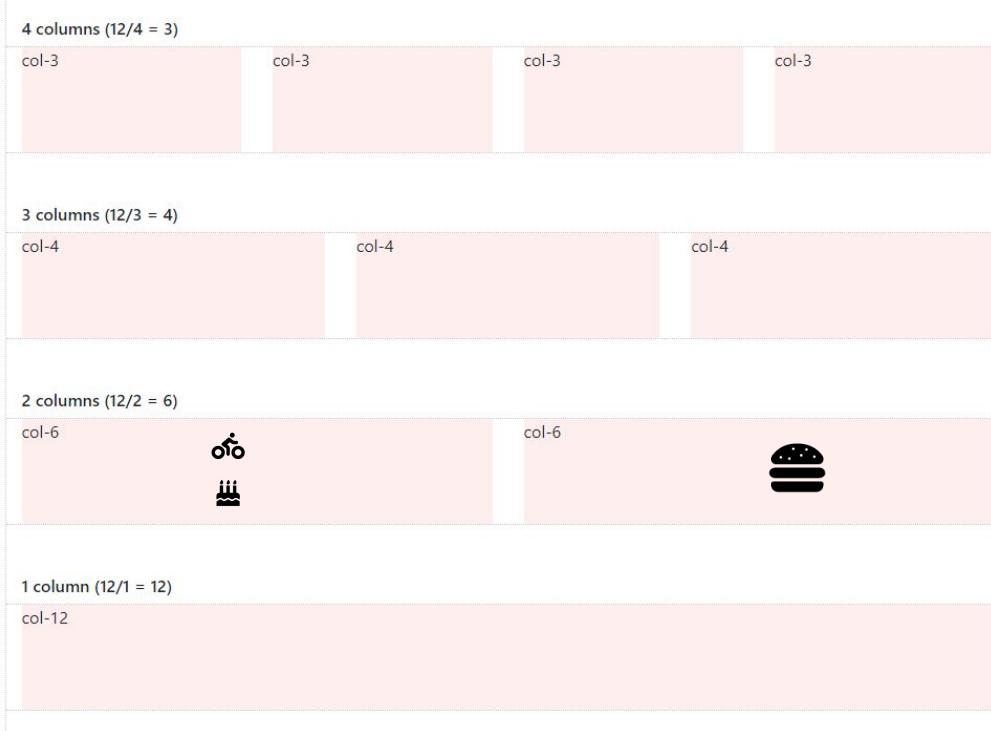
How do we create this?

```
fluidRow(  
  column(3, sele  
  ctInput()),  
  column(9, plotOutput())  
)
```



**How do you stack things
in the same row?**

Use div() for rows!



```
fluidRow(  
  column(6,  
    div({ofo}, {cake}))  
  
  column(6, {burger})  
)
```

[Image source](#)

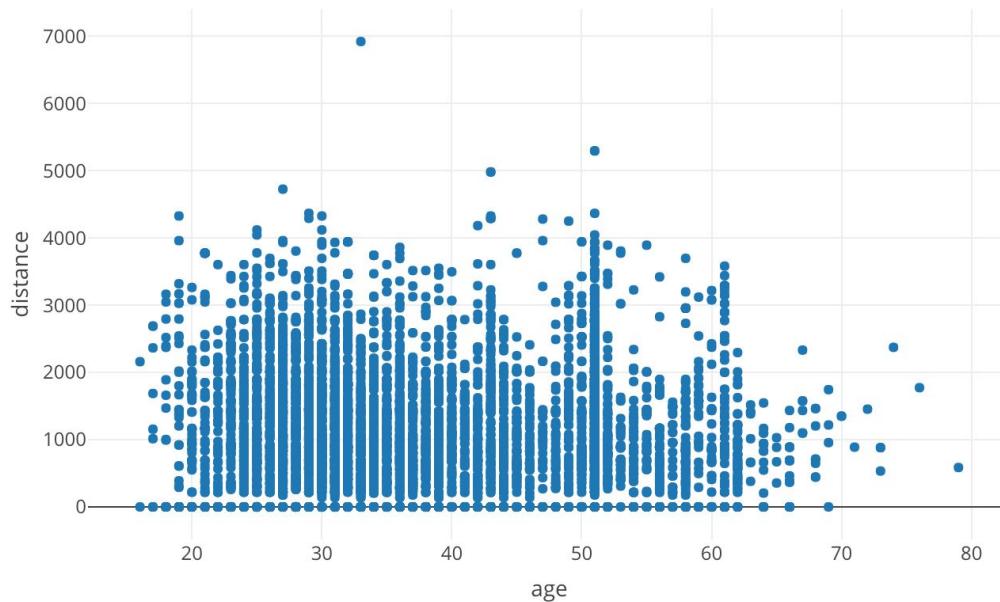
Can you add paragraph text here?

col-3

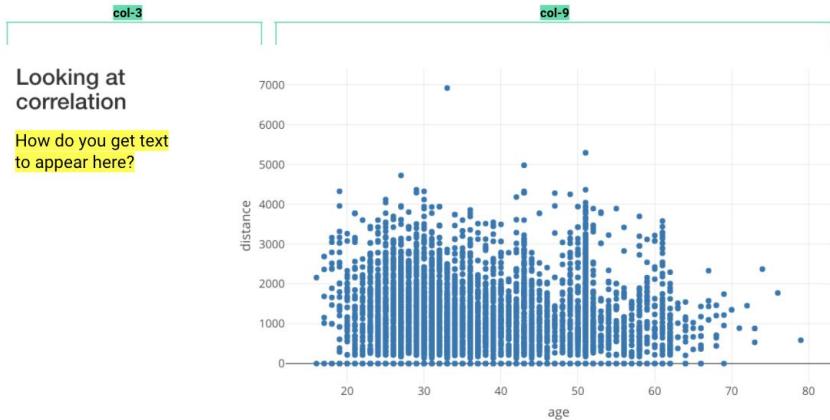
col-9

Looking at correlation

How do you get text to appear here?

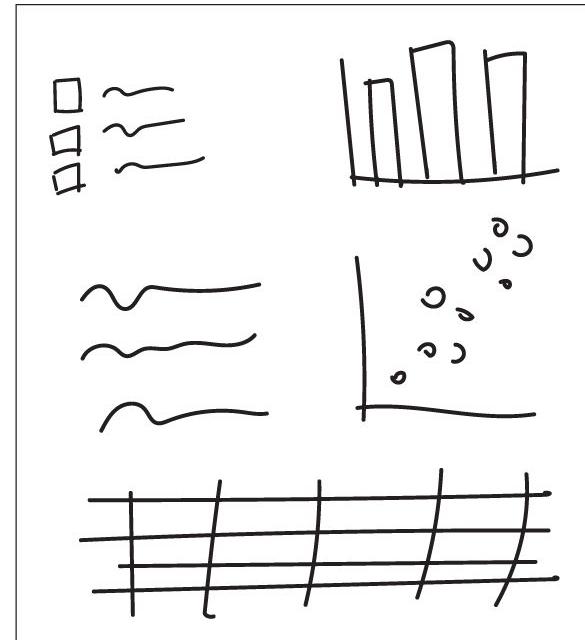
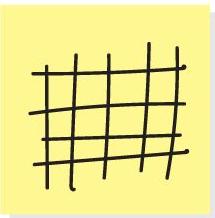
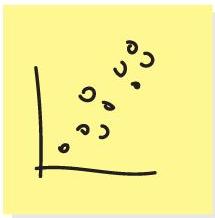
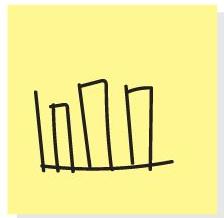


Can you add an action button here?

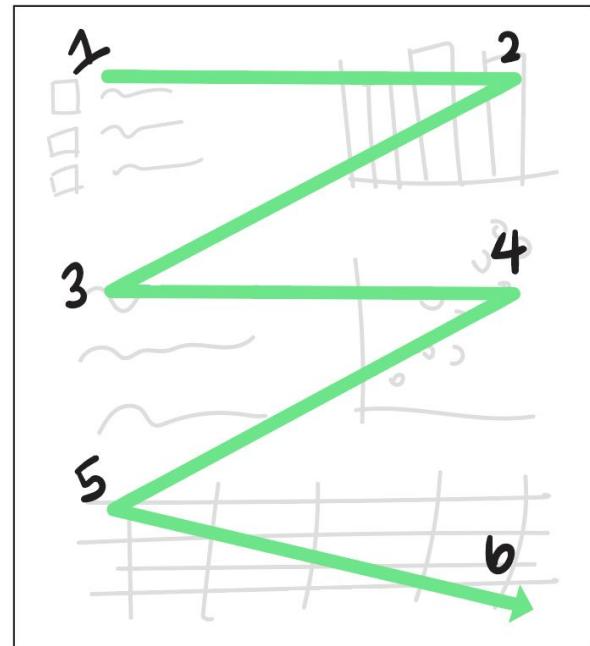


```
fluidRow(  
  column(3, h3())  
) ,  
  
  column(9, plotOutput())  
)
```

Start on Paper!

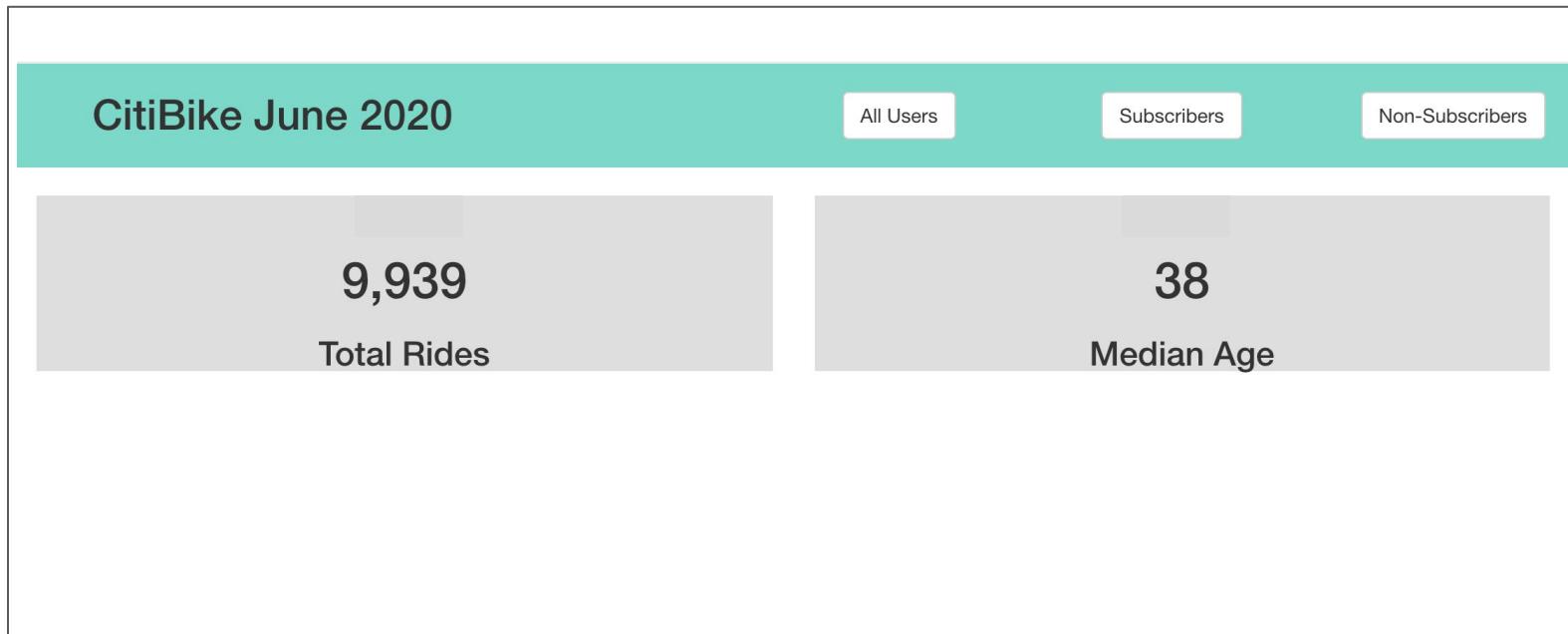


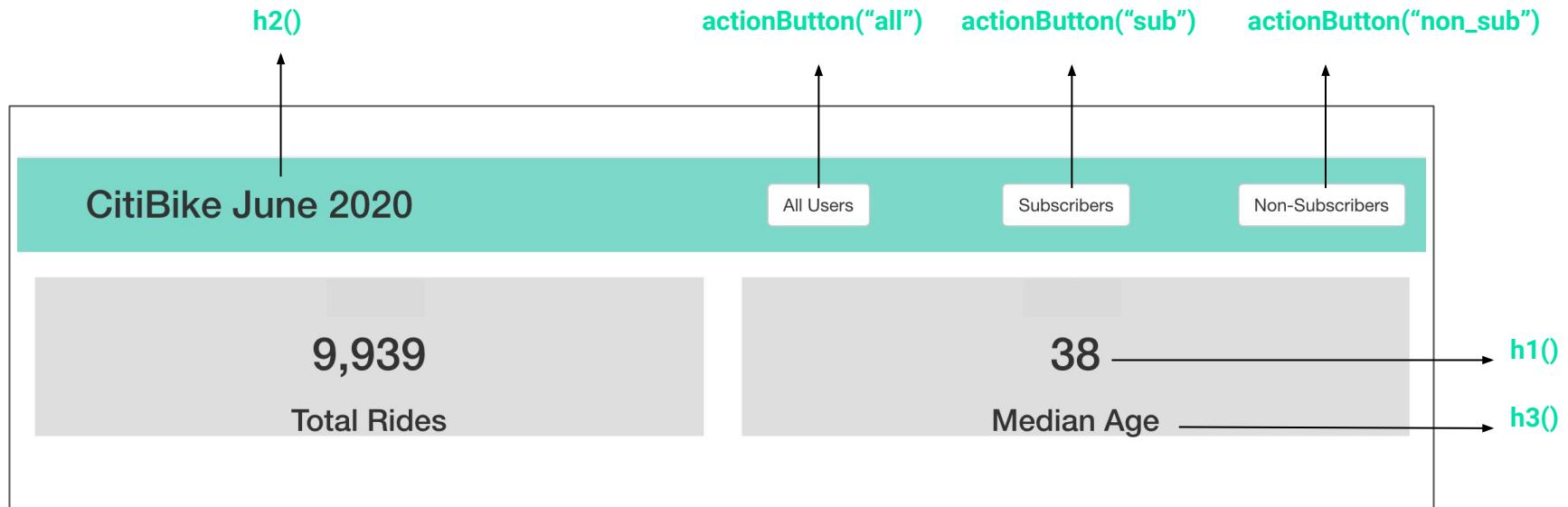
**Keep in mind the
direction of how we read!**



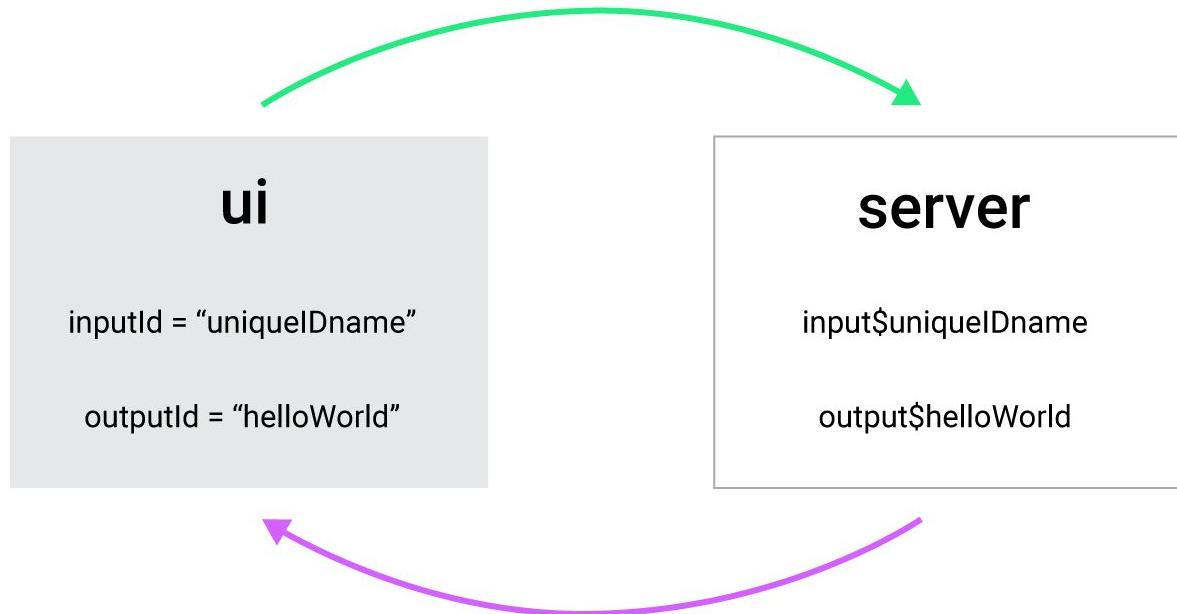


**Draw this layout on paper.
Label each UI elements, col width, div, etc..**





open 03_final/layout.R



CitiBike June 2020

All Users

Subscribers

Non-Subscribers

9,939

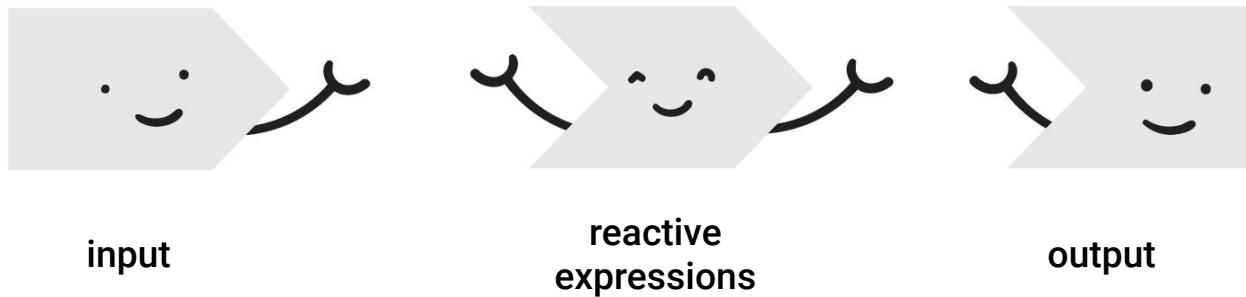
Total Rides

38

Median Age

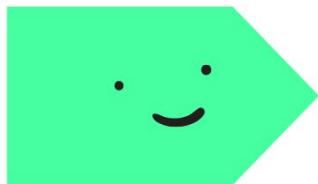
Exercise: Create
this value

Reactive Players

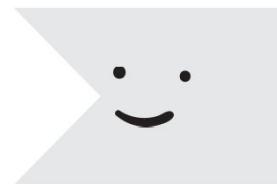


[using shapes from reactlog](#)

Basic Reactive Flow



input



output

- Passes input variable to be used in output
- No new object created
- No variables changed

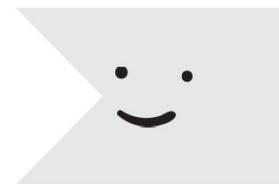
More Complex Reactive Flow



input

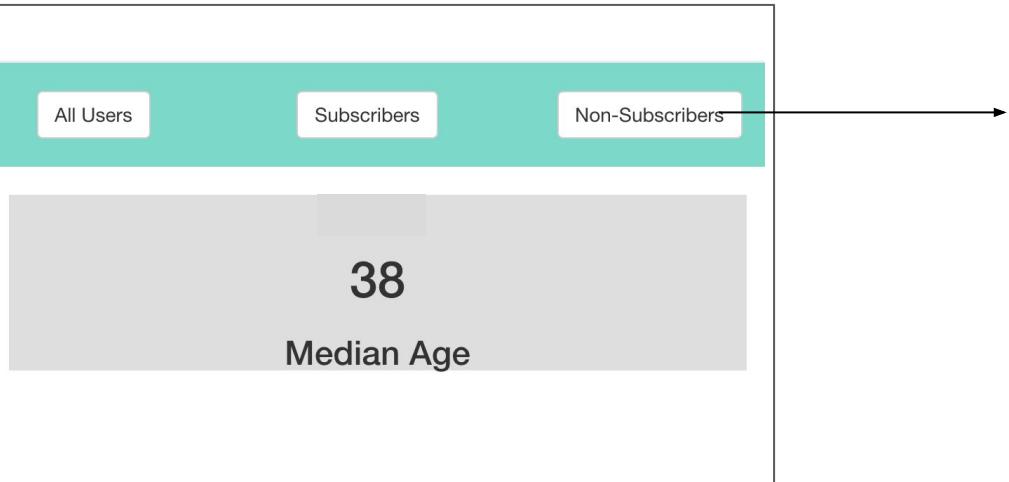


reactive
expressions



output

- Expressions can store reactive values
- Can change variables
- Tracks dependencies



What do we want
this button to do?

1. Filter data based on non-subscribers
2. Run all the code on the page with filtered data

open 03_final/layout.R



9,939

Total Rides

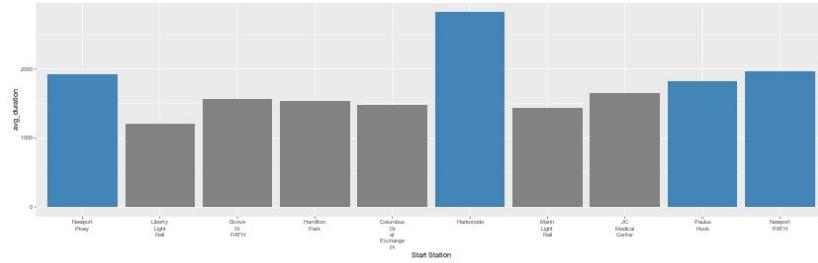


38

Median Age

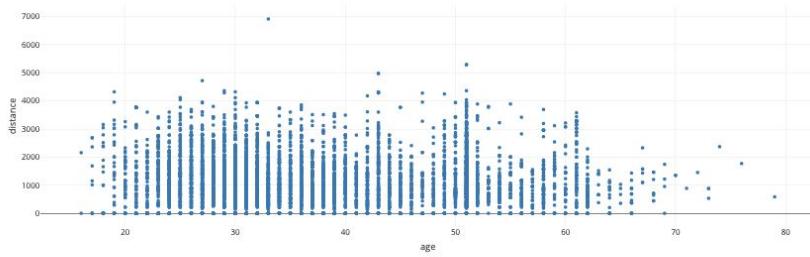
Select Variable

Avg Duration



Looking at correlation

This is an example of using plotly and show how tooltips work.



	start_station_name	total_rides	avg_duration	avg_age
1	Newport Pkwy	742	1822	38
2	Liberty Light Rail	676	2821	39
3	Grove St PATH	500	1200	39
4	Hamilton Park	438	1559	40
5	Columbus Dr at Exchange Pl	427	1929	38
6	Harborside	419	1541	38
7	Marin Light Rail	394	1434	38
-	-	-	-	-

Shiny App

Look in the '04_final'
directory for how to put it
all together to create this!

Key Takeaways

1. Always start on paper
2. Invest time to understanding your audience
3. Try everything in regular R code before bringing it into Shiny
4. Choose your visual cues wisely
5. Design for happiness :)

Extra goodies

Check out 'shiny_dashboard' folder

VIVIAN **PENG**

Thank you!

<https://www.vivianpeng.com>