## Data Wrangling and Data Analysis Data Visualization (Creating Dashboards)

### **Hakim Qahtan**

Department of Information and Computing Sciences
Utrecht University



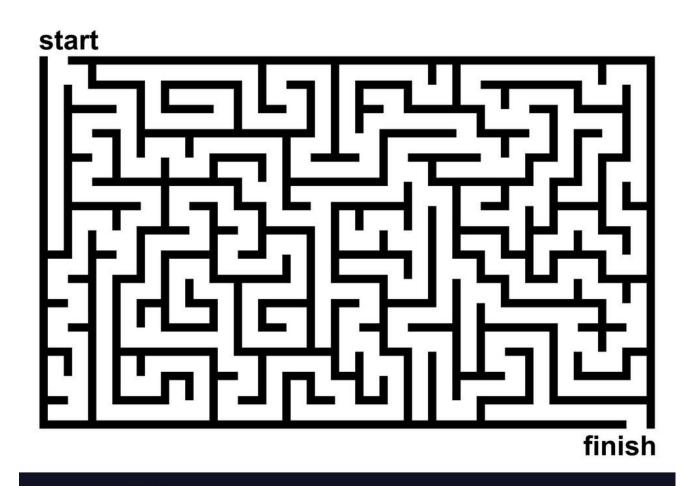
### **Topics for Today**

- Introduction
- Mistakes in Data Visualization
- Examples of Visualization Tools for Data Analytics
- Different Ways to Represent Data
- Designing a Dashboard



### Introduction

- Danger of getting lost in data, which may be:
  - Irrelevant to the current task in hand
  - Processed in an inappropriate way
  - Presented in an inappropriate way



VectorStock®

VectorStock.com/24738840

# Introduction (Cont.)

#### **VISUAL-ANALYTICS.EU**

the new visual analytics portal

What is Visual Analytics?

Book ▼

Related Websites Contact

News Conferences iVAC In the Media Interactive Demos Job Offers Publications

#### **FEATURED POSTS**



#### EuroVA 2014, Swansea, June 9-10

EuroVA 2014 is the fifth international Eurovis workshop on Visual Analytics held in Europe and aims at continuing the success of the previous editions, held in Bordeaux, France, in Bergen, Norway, in Vienna, Austria and in Leipzig, Germany. Also this [...]



#### "MASTERING THE INFORMATION AGE" now on visual-analytics.eu

Mastering The Information Age - Solving Problems with Visual Analytics

The VisMaster book is now available in print form but you can also still download the full book from here [25 MB, PDF].

Go to the Book menu on the top bar for a summary of the book and individual chapters.

(Note that individual chapters can be downloaded separately if you wish)





#### Workshop and special issue on 'Progress in Movement Analysis - Experiences with Real Data'

Organisers and Guest Editors of 'Computers, Environment and Urban Systems' (CEUS)...



#### The Second UKVAC International Visual Analytic Summer School (VASS2012)

The Second UKVAC International Visual Analytic Summer School (VASS2012) Date: 3-8^th September...



#### Call for Papers: SIGRAD 2012 -Interactive Visual Analysis of

Call for Papers: SIGRAD 2012 -Interactive Visual Analysis of Data Location: Växjö, Sweden Date:..



#### **TAProViz Call for Papers**

Visualizations can make the structure and dependencies between elements in processes accessible...

#### IN THE MEDIA More in In the Media

Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats - BBC Four

Hans Rosling's famous lectures combine enormous quantities of public data with a sport's commentator's style to reveal the...



Video Interview - JD Fekete talks about Jacques Bertin

Journalism in the Age of Data: Visualization as a Storytelling Medium

David McCandless: The beauty of data visualization

LOGIN
Username
Password
Login →

### **Introduction (Cont.)**

- Good graphics
  - Point relationships, trends or patterns
  - Explore data to infer new things
  - To make something easy to understand
  - To observe a reality from different viewpoints
  - To achieve an idea to be memorized

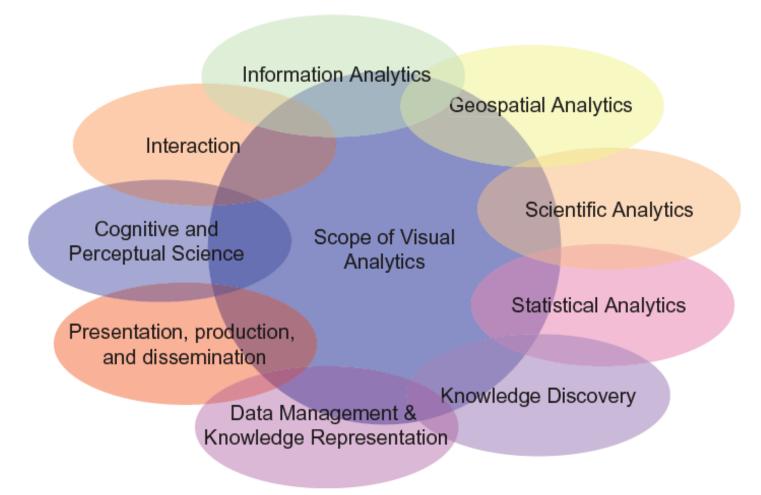


### **Introduction (Cont.)**

- Data visualization is a way of expressing
  - Like math, music, drawing or writing
- So, it has some rules to respect

$$\begin{array}{l} = \frac{n!}{k! \cdot (n-k)!} + \frac{n!}{(k+1)! \cdot (n-(k+1))} \\ = \frac{(k+1)! \cdot b!}{(k+1) \cdot k! \cdot (n-k)!} + \frac{n! \cdot (n-k)!}{(k+1)! \cdot (n-k)!} \\ = \frac{(k+1)! \cdot n!}{(k+1)! \cdot (n-k)!} + \frac{n! \cdot (n-k)!}{(k+1)! \cdot (n-k)!} \\ = \frac{(k+1)! \cdot n!}{(k+1)! \cdot (n-k)!} + \frac{n! \cdot (n-k)!}{(k+1)! \cdot (n-k)!} \\ = \frac{(k+1)! \cdot (n-k)!}{(k+1)! \cdot (n-k)!} + \frac{n! \cdot (n-k)!}{(k+1)! \cdot (n-k)!} \\ = \frac{(k+1)! \cdot (n-k)!}{(k+1)! \cdot (n-k)!} + \frac{n!}{n!} \\ = \frac{(k+1)! \cdot (n-k)!}{(k+1)! \cdot (n-k)!} \\ = \frac{(n-k)!}{(k+1)!} + \frac{n!}{n!} \\ = \frac{n!}{(2n-1)!} \\ = \frac{(n+1)! \cdot (n-k)!}{(n+1)!} \\ = \frac{(n+1)! \cdot (n-k)!}{(n-k)!} \\$$

### **Introduction (Cont.)**



Keim, D.A., Mansmann, F., Schneidewind, J. and Ziegler, H., Challenges in Visual Data Analysis, Tenth International Conference on Information Visualisation (IV'06), pp. 9-16, 2006.



### **Mistakes in Data Visualization**





If it's an image that displays and explains information quickly and clearly, it's an infographic. But we've collected some that are head-craning, eye-squinting, eyebrow-raising nightmares that leave you more confused than before you clicked 'next'. The result is an exciting gallery of infographics that tell you nothing.

If you manage to understand all 16 of the infographics in this gallery, write to data@theguardian.com and we'll pick one lucky person to win our incredible prize (the prize is a ruler and a ball-point pen).

#### Mona Chalabi

Thu 1 Aug 2013 12.55 BST

It's quite ironic that you need a map to navigate this map of the US. The key in the bottom-left corner explains that the colours are based on three maps overlayed. Oh! Right! We get it now! (We don't really, it's just embarrassing to admit sometimes).

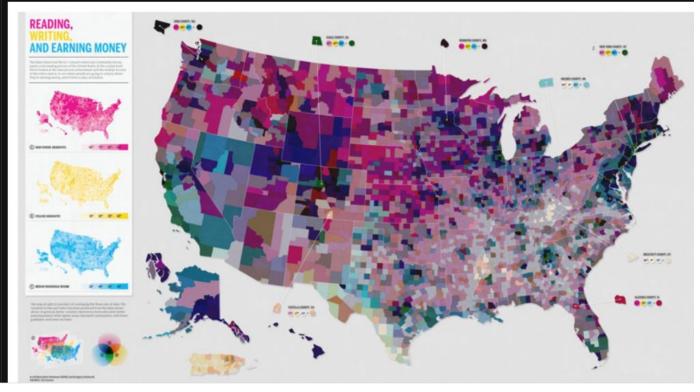
Photograph: GOOD/Gregory Hubacek





f





### Mistakes in Data Visualization (Cont.)



• Problems?



### Mistakes in Data Visualization (Cont.)

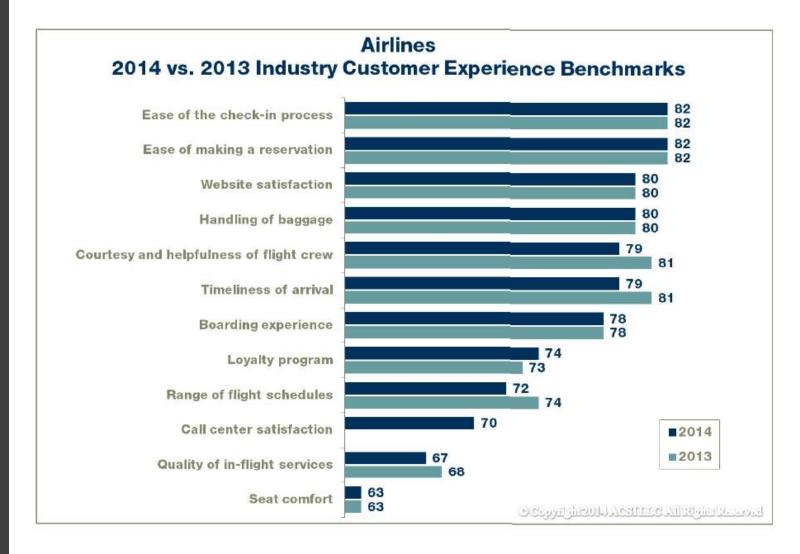
- Problems?
- Try to identify:
  - 1. The biggest donor in 2008
  - 2. The smallest donor in 2009
  - 3. The variation between 200 and 2009
  - 4. Which region received the biggest amount of money



https://blogs.elpais.com/.a/6a00d8341bfb1653ef0153903125d9970b-550wi

### Mistakes in Data Visualization (Cont.)

- Problems?
  - Bar values should start at zero



https://www.qualitydigest.com/IQedit/Images/Articles and Columns/2014/ April 2014/ACSI/airlines-lg.jpg

### **Tools for Data Analytics**



### Requirements for DA Visualization Tools

- Ability to choose different visuals and graphs
  - Choose between pie charts, bar charts, and line graphs to express the data better
- Accurate trend tracking capability
  - Keep track of the interesting trends in your data
  - Ex.: you can see which products are selling each month, and which demographic is buying
- High level of security
  - Without proper security, you could risk losing information that could be crucial to your enterprise
- Simplified and self-explainable software interfacing



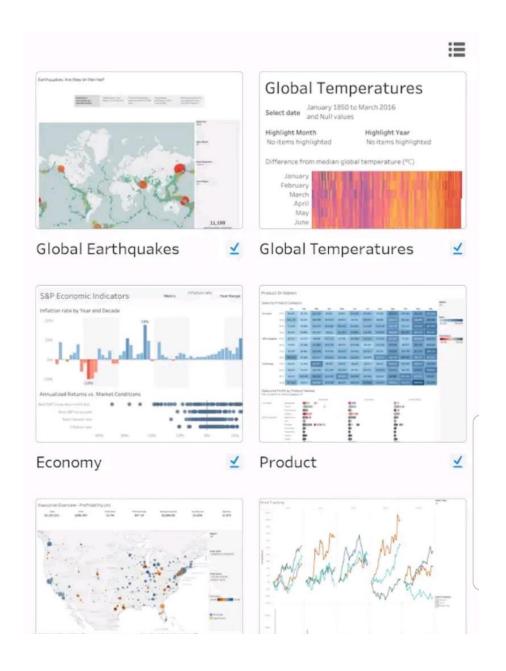
### **Tableau**

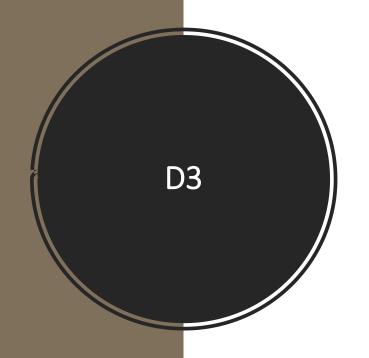
- Connect to and aggregate data.
  - Flat files, databases
  - Perform joining and clustering
- Create insightful analysis.
  - Explore data, create good looking charts and dashboards online
  - Drag and drop functionality
  - Tell compelling stories with your data
  - Write custom calculations within your visualizations
  - Set up filters

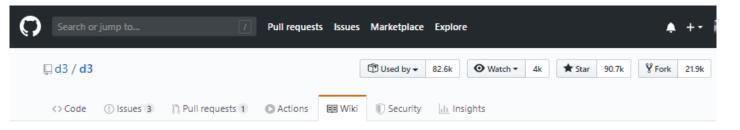


### Tableau (Cont.)

- Free for one data source which should be made public.
- Requires purchase of a license for private use
  - 70 USD per month.







#### Gallery

Mike Bostock edited this page 26 days ago · 1295 revisions

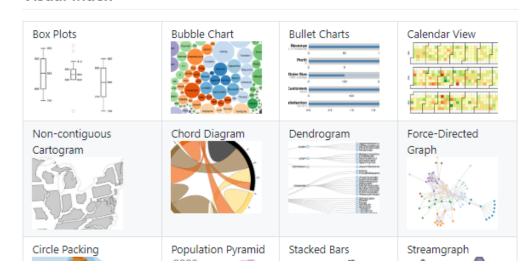
Wiki ► Gallery

#### Please see the new D3 Gallery on Observable.

Many examples below are broken or use older versions of D3. We're in the process of cleaning up this gallery, and writing new examples and tutorials. Thanks for your patience!

Welcome to the D3 gallery! More examples are available for forking on Observable; see D3's profile and the visualization collection. Please share your work on Observable, or tweet us a link!

#### **Visual Index**





### Data-Driven Documents

Home

▶ Pages 62

- Introduction
- Examples
- Tutorials
- Plugins

#### Help

- Stack Overflow
- Slack (Invite)
- Google Group
- Gitter

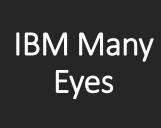


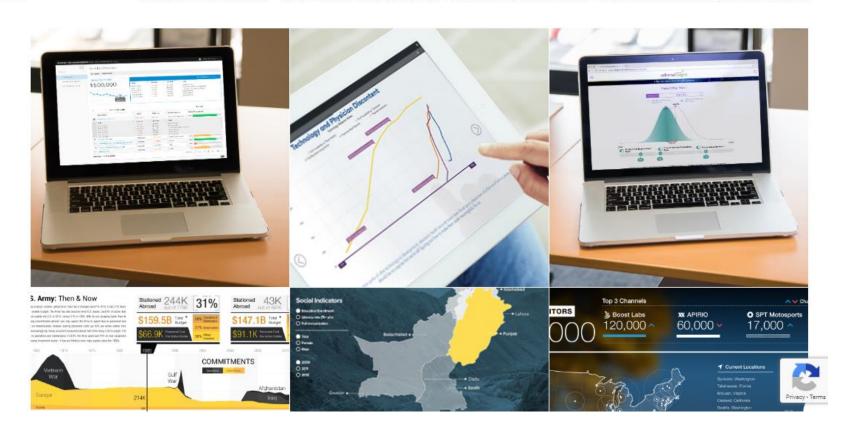


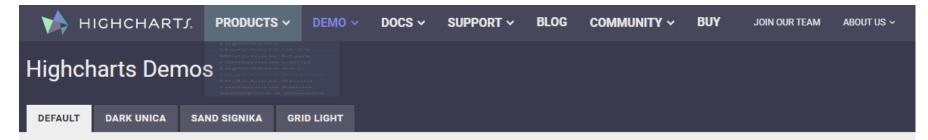


SHOW ALL

DATA ANALYTICS SOLUTIONS DATA VISUALIZATION & REPORTING DATA PRODUCT DEVELOPMENT DATA PRODUCT UI/UX DESIGN & DEVELOPMENT

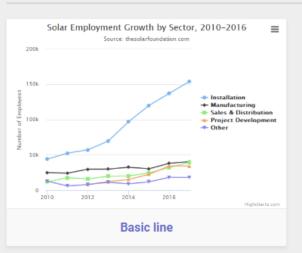




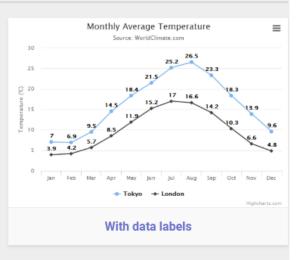


#### Line charts

Highcharts

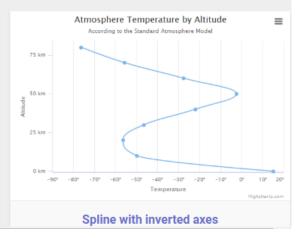












Support Home Guides Reference Send fee

Overview

Hello, Charts!

Ouickstart

Load the Charts Library

Prepare the Data

Customize the Chart

Draw the Chart

Draw Multiple Charts

Chart Types

**Chart Gallery** 

Annotation Charts

Area Charts

Google Charts

Bar Charts

**Bubble Charts** 

Calendar Charts

Candlestick Charts

Column Charts

Combo Charts

Diff Charts

Donut Charts

Gantt Charts

Gauge Charts

GeoCharts

Histograms

Intervals

Line Charts

Maps

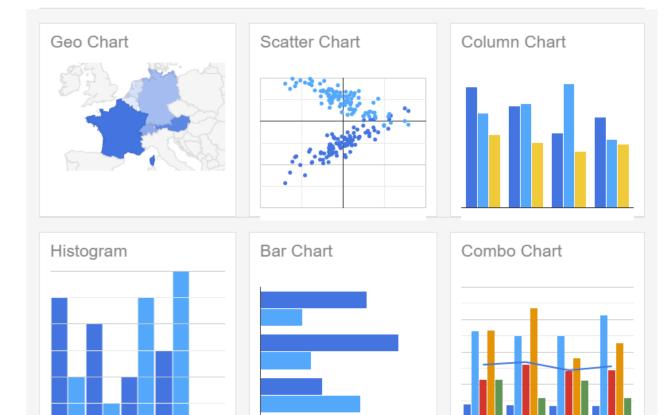
Org Charts

Pie Charts

#### Chart Gallery

Our gallery provides a variety of charts designed to address your data visualization needs. These charts are based on pure HTML5/SVG technology (adopting VML for old IE versions), so no plugins are required. All of them are interactive, and many are pannable and zoomable. Adding these charts to your page can be done in a few simple steps.

Some additional community-contributed charts can be found on the Additional Charts page.



#### NodeXL

download archive

NodeXL: Network Overview, Discovery and Exploration for Excel

NodeXL is an Excel 2007/2010/2013 template for viewing network graphs, with a set of .NET Framework 3.5 class libraries that can be used to add network graphs

> issues discussions home



Donate to the Social Media Research Foundation's NodeXL Development Fund

NodeXL is the MSPaint of Networks.

Excel Window

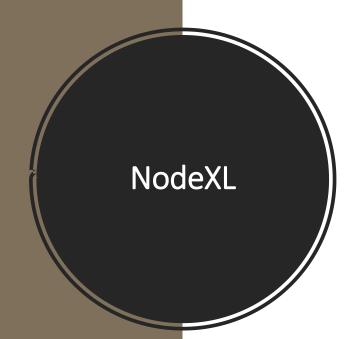
NodeXL Basic is a free, open-source template for Microsoft® Excel® 2007, 2010, 2013 and 2016 that makes it easy to explore network graphs. With NodeXL, you can enter a network edge list in a worksheet, click a button and see your graph, all in the familiar environment of the Excel window.

NodeXL Pro offers additional features that extend NodeXL Basic, providing easy access to social media network data streams, advanced network metrics, and text and sentiment analysis, and powerful report generation. NodeXL Pro can create insights into social media streams with just a few clicks.

NodeXL Pro released on October 12, 2015.



Visit the NodeXL Graph Gallery to see the wide variety of graphs that have been created by the NodeXL community.



#### What is Power BI?

Power BI is a business analytics solution that lets you visualize your data and share insights across your organization, or embed them in your app or website. Connect to hundreds of data sources and bring your data to life with live dashboards and reports.

WATCH OVERVIEW >

WATCH DEMO >



### Microsoft Power BI

- Unlimited access to on-site and in-cloud data to give a centralized data access hub
- Comes with a robust tutorial system to help in learning how to use it properly
- Allows to import and publish data online for your collaborators
- Free and 10\$/month for pro-package



### **More DA Visualization Tools to Consider**

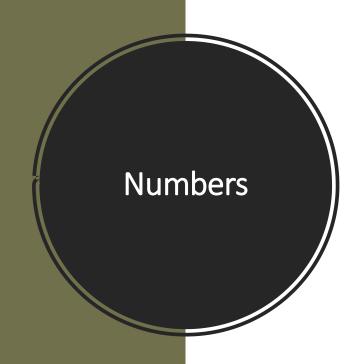
- IBM Watson Analytics
- Infogram
- ChartBlocks
- Datawrapper
- Sisense
- Looker
- Periscope Data
- Zoho Analytics

- Domo
- Qlikview
- Klipfolio
- SAP Analytics Cloud
- Kibana
- Plotly
- Chartio
- •



### **Choose How to Represent Your Data**







https://www.paulkrekorian.org/los\_angeles\_tourism\_on\_the\_rise

### **Tables**

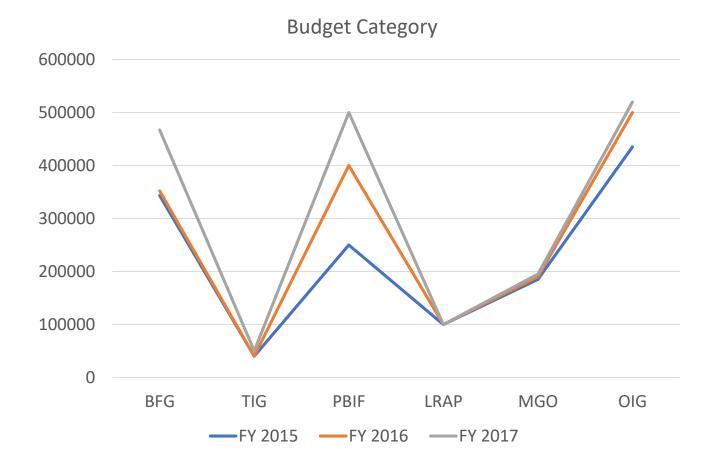
- Could be helpful when talking with sophisticated audience
- Also, when is the data is reasonably small

Budget Category	FY 2015	FY 2016	FY 2017
BFG	343150	352000	467000
TIG	40000	40000	50000
PBIF	250000	400000	500000
LRAP	100000	100000	100000
MGO	185000	190000	195000
OIG	435000	500000	520000



### Lines

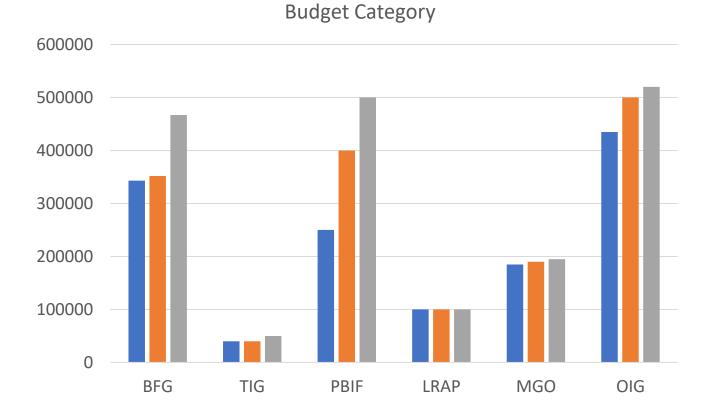
• Likely, best way to show trends in the data





### Bars

- Most popular way of presenting data
- Shows insights of the data that is not suitable for line charts



FY 2016

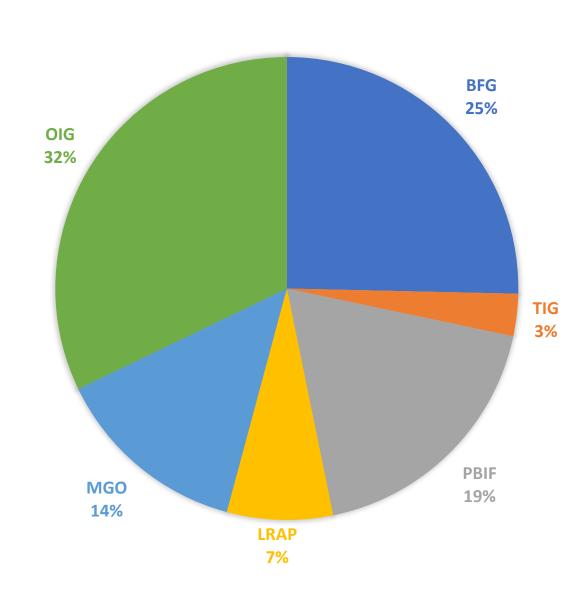
FY 2015

■ FY 2017



### Pie Charts

- Nothing pie charts can do better than bar charts
- Could be reasonable to highlight two or three percentages among the rest

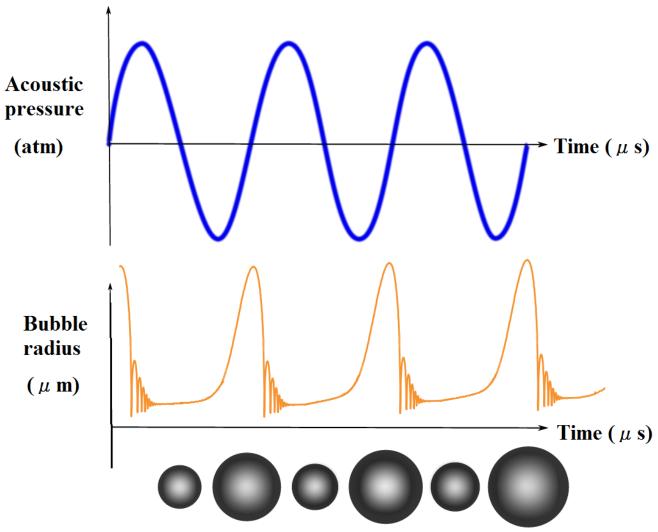


FY 2015



### **Bubble Plot**

Scatter (or bubble)
 plots can show the
 trends of a lot of
 different data points

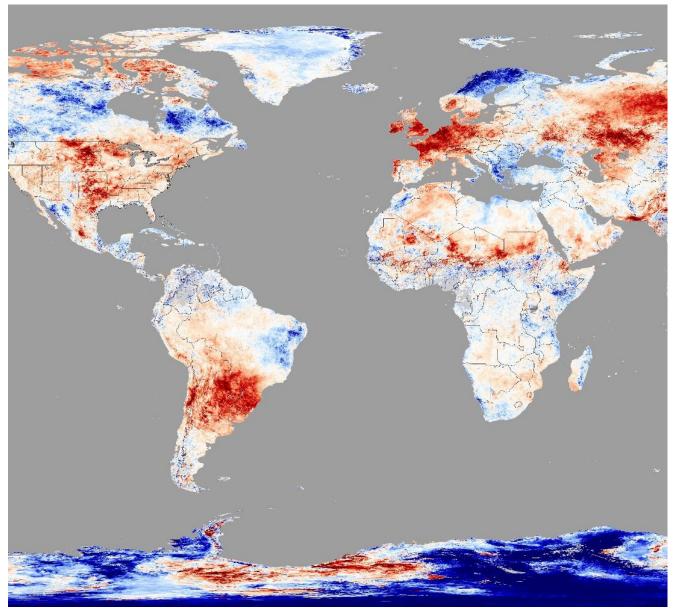


https://commons.wikimedia.org/wiki/File:Acoustic pressure vs bubble radius.png



### Map Plot

- Maps can be a powerful way to represent data geographically
- In the figure, land surface temperature anomaly (-10  $:10^{o}$ C)



https://eoimages.gsfc.nasa.gov/images/imagerecords/6000/6803/globallstanom\_t mo\_2006193\_lrg.jpg



### **Designing a Dashboard**



### **Dashboards**

### **Fundamentals**

Perception
Vision
Color
Principles

### **Techniques**

Representation Presentation Interaction

### **Application**

Dashboard Visual Analytics



### **Dashboards** (Cont.)

"Most information dashboards that are used in business today fall far short of their potential"

Stephen Few



### **Dashboards – Definition**

"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance"

Few2007



### **Dashboards – Characteristics**

- Visual displays
- Display information needed to achieve specific objectives
- Fits on a single computer screen
- Are used to monitor information at a glance
- Have small, concise, clear, intuitive display mechanisms
- Are customized



### **Dashboards – Common Mistakes**

- Supplying inadequate context for the data
  - Fail to provide adequate context to make the measures meaningful
- Displaying excessive detail or precision
  - Show unnecessary detail
- Choosing a deficient measure
  - Use of measures that fail to directly express the intended message
- Choosing inappropriate display media
  - Common problem with pie charts
- Introducing meaningless variety
  - Exhibit unnecessary variety of display media



### **Dashboards – Common Mistakes (Cont.)**

- Using poorly designed display media
  - A legend was used to label and assign values to the slices of the pie. This forces our eyes to bounce back and forth between the graph and the legend to glean meaning, which is a waste of time and effort when the slices could have been labeled directly.
  - The order of the slices and the corresponding labels appears random. Ordering them by size would have provided useful information that could have been assimilated instantly.
  - The bright colors of the pie slices produce sensory overkill. Bright colors ought to be reserved for specific data that should stand out from the rest.
- Encoding quantitative data inaccurately



### **Dashboards – Common Mistakes (Cont.)**

- Arranging the data poorly
  - The most important data ought to be prominent
  - Data that require immediate attention ought to stand out
  - Data that should be compared ought to be arranged and visually designed to encourage comparisons
- Highlighting important data ineffectively or not at all
  - Fail to differentiate data by its importance
    - Giving relatively equal prominence to everything on the screen
- Cluttering the display with useless decoration
  - Try to look something that is not
  - It results in useless and distracting decoration

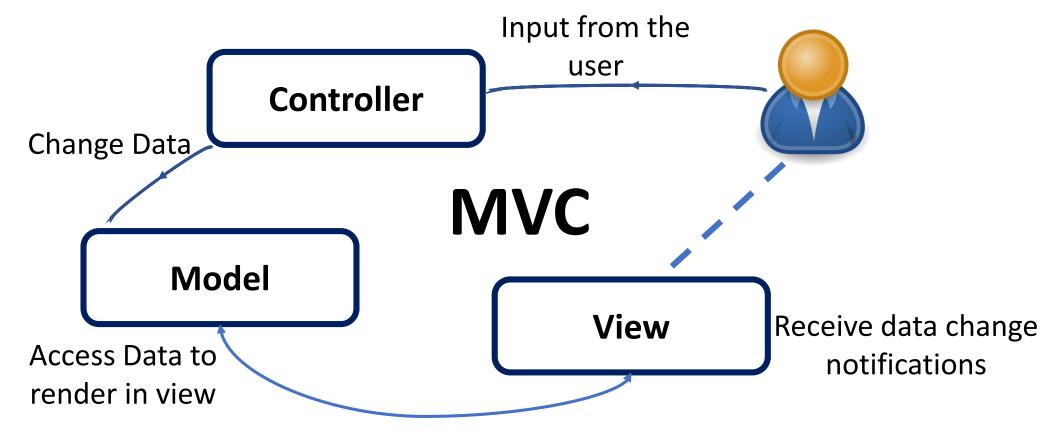


### **Dashboards – Common Mistakes (Cont.)**

- Misusing or overusing color
  - Too much color undermines its power
- Designing an unattractive visual display
  - The fundamental challenge of dashboard design is to effectively display a great deal of often disparate data in a small amount of space



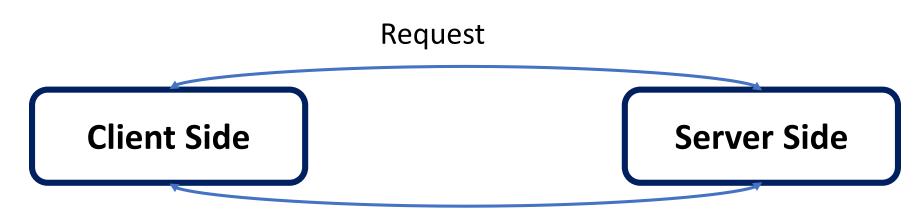
### **Model-View-Control**



"The controller is essential and explicit: you have to specify what to do when you receive user requests and what resources you are going to mobilize to carry out the necessary tasks outlined in the model"



#### **Reactive Web Framework**



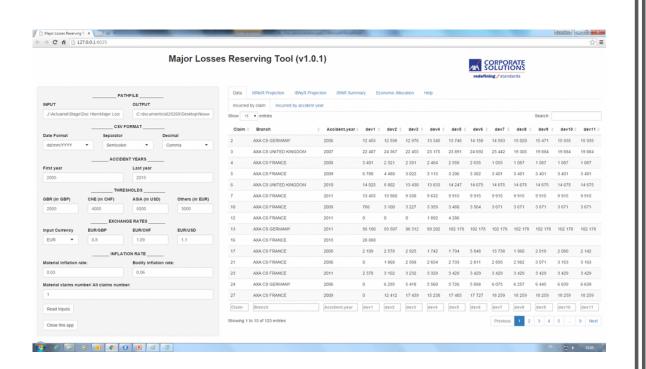
Response

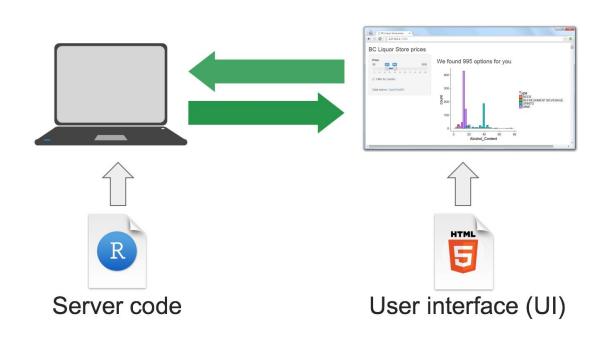
"Reactive Systems are highly responsive, giving users effective interactive feedback"

<u>https://www.reactivemanifesto.org</u>
<u>http://littleactuary.github.io/blog/Web-application-framework-with-Shiny/</u>



# Reactive Web Framework



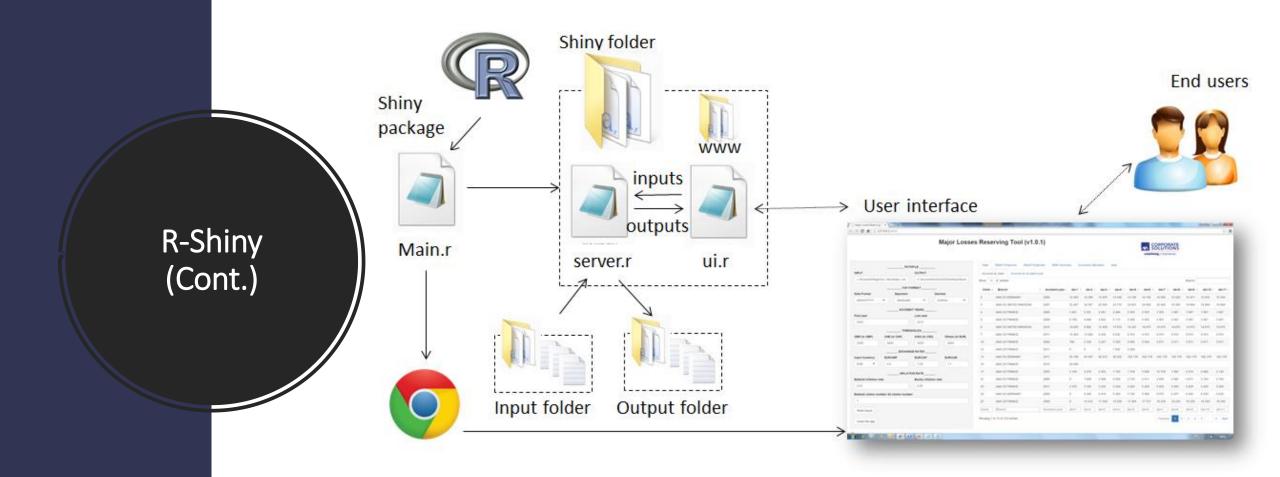




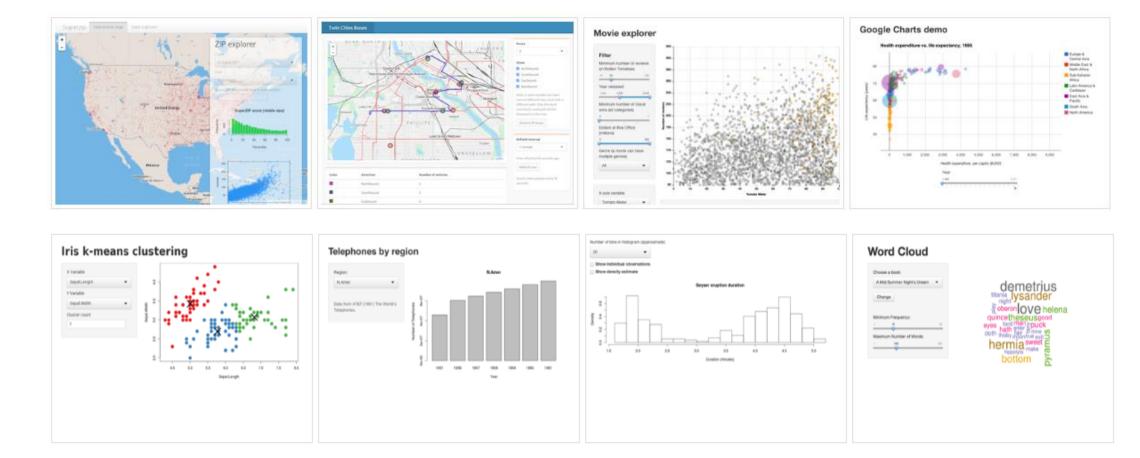
### **R-Shiny**

- What is Shiny?
  - R package that allows programmers to:
    - Quickly build a user interface to explore and work with data.
    - Publish their charts and graphs online.
- Why Shiny?
  - Interactive display and manipulation of data
  - Interactive communication sessions between the user's browser and a server without having to poll the server for a reply
  - Easy to develop and share with clients and project teams
  - Entirely extensible custom input/output
  - Open source library





Wide variety of templates that you can get inspired by





- Shiny application consists of two main components
  - The user interface (frontend) component
    - Allows the user to interact with the application
  - The server (backend) component:
    - Listens to the requests of the user and respond accordingly based on predefined set of functionalities
- A nice tutorial can be found at

https://shiny.rstudio.com/tutorial/



- Shiny application consists of two main components
  - The user interface (frontend) component
    - Allows the user to interact with the application
  - The server (backend) component:
    - Listens to the requests of the user and respond accordingly based on predefined set of functionalities
- A nice tutorial can be found at

https://shiny.rstudio.com/tutorial/



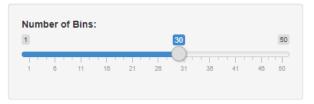
# Two Ways to Create Shiny Application

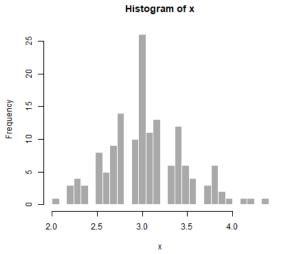
- 1. Use 2 R script files ui.R and server.R Rstudio has made it easy for you to incorporate widgets and charts (ui side) and run algorithms on your data (server.R)
  - Click on 
    Run App 
    or type runAPP() in the console.
- 2. Use app.R and store the user interface widgets and charts in one object (let's say *u*) and the algorithms in another object (let's say *s*)
  - Call shniyApp(ui = u, server = s)



# ui.R

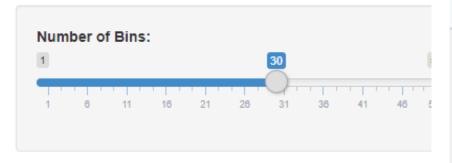
#### Histogram Example

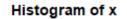


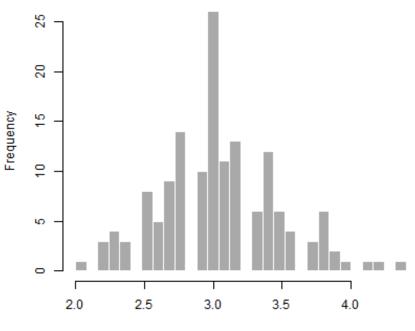


```
ui.R × Server.R ×
    | 🖅 | 🔒 | 🔍 🎢 🗸 📗
     library(shinydashboard)
     fluidPage(
       # Application Title
       titlePanel('Histogram Example'),
       # Sidebar with a slider input for the number of bins
       sidebarLayout(
         sidebarPanel(
           sliderInput('bins', 'Number of Bins:',
                       min = 1, max = 50, value = 30),
 10
         mainPanel(
 11
           plotOutput('histPlot')
 12
 13
 14
 15
```

### Histogram Example

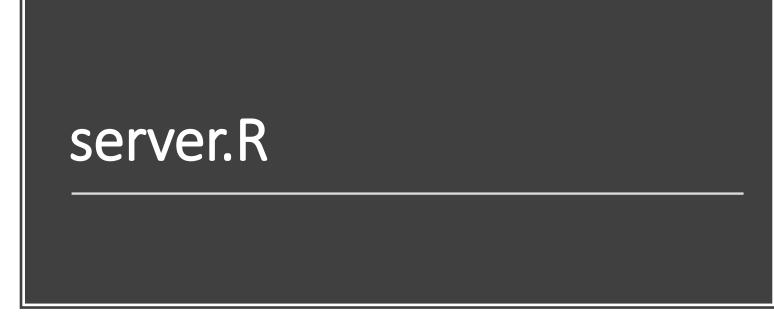




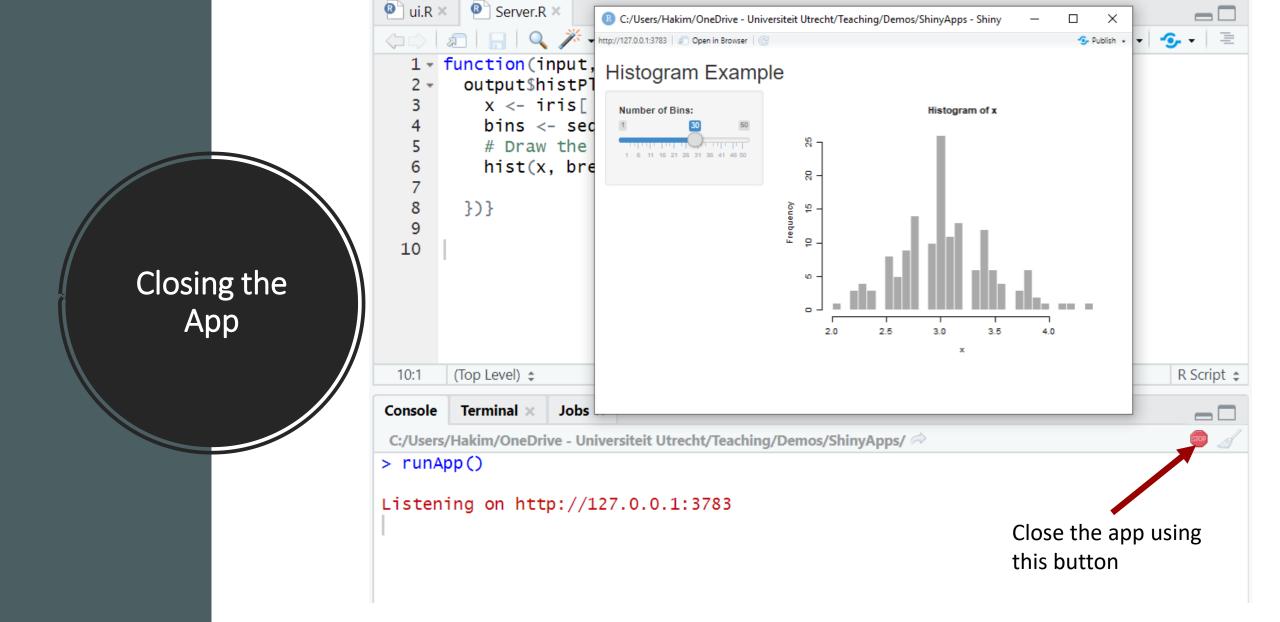


```
iui.R * Server.R *

1 function(input, output, session) {
    output$histPlot <- renderPlot({
        x <- iris[ , 2]
        bins <- seq(min(x), max(x), length.out = inj
        # Draw the histogram using the specified number hist(x, breaks = bins, col = 'darkgray', bound to be a served.)
}
</pre>
```





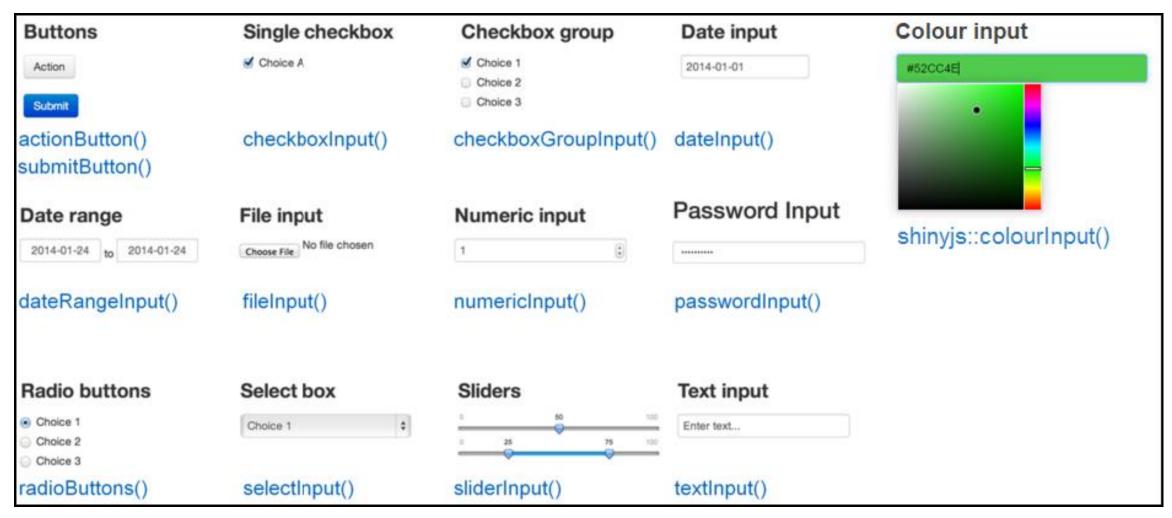


# **Application Template**

```
library(shinydashboard)
# or library(shiny) for simple visualizations
u <- dashboardPage(dashboardHeader(),
      dashboardSidebar(), dashboardBody())
# dashboardPage expects three arguments
# You can also use pageWithSidebar() or fluidPage() ...
s <- function(input, output){}
shinyApp(ui = u, server = s)
What is the output of this app?
```



## **R-Shiny – Basic Widgets**



Example: numericInput(inputId = "", label = "", min = 0, max = 10, step = 1, value = 5, width = 200)



- Shiny application consists of two main components
  - The user interface (frontend) component
    - Allows the user to interact with the application
  - The server (backend) component:
    - Listens to the requests of the user and respond accordingly based on predefined set of functionalities
- A nice tutorial can be found at

https://shiny.rstudio.com/tutorial/



### **R-Shiny and HTML**

- How the parts of a dashboard work together?
- HTML tag functions in Shiny, like div() and p() return objects that can be rendered as HTML

```
# A basic div
div(class = "my-class", "Div content")
## <div class="my-class">Div content</div>
# Nested HTML tags
div(class = "my-class", p("Paragraph text"))
## <div class="my-class">
## Paragraph text
## </div>
```



### R-Shiny and HTML (Cont.)

• Some functions in Shiny, returns more complex HTML fragments

```
textInput("Id", "Label")
                                          sidebarPanel(
                                            div("First div"),
## <div class="form-group shiny-
input-container">
                                            div("Second div")
     <1abel for = "Id">
Label</label>
                                          \#\# < \text{div class} = \text{"col-sm-4"} >
     <input id="Id" type="text"</pre>
                                          ##
                                              <form class="well">
class="form-control" value=""/>
                                          ##
                                                 <div>First div</div>
## </div>
                                          ##
                                                 <div>Second div</div>
                                               </form>
                                          ## </div>
```



### R-Shiny and HTML (Cont.)

• To check the equivalent HTML of shiny functions, run the shiny function in the r-console

```
Terminal ×
Console
                 Jobs ×
C:/Users/Hakim/Desktop/
> selectInput('xcol', 'X Variable', names(iris))
<div class="form-group shiny-input-container">
  <label class="control-label" for="xcol">X Variable</label>
  <div>
    <select id="xcol"><option value="Sepal.Length" selected>Sepal.Length
<option value="Sepal.Width">Sepal.Width</option>
<option value="Petal.Length">Petal.Length
<option value="Petal.Width">Petal.Width</option>
<option value="Species">Species</option></select>
    <script type="application/json" data-for="xcol" data-nonempty="">{}</script>
  </div>
</div>
```



### **Demo**

- In the demo, you will learn how to:
  - Create simple interactive plot using Shiny
  - Select parameters using selectInput() and numericInput()
  - Select input file and reading them
  - Rend plots
  - Create conditional panels



## **Reading Material**

- Useful resources:
  - Shiny official tutorial -http://shiny.rstudio.com/tutorial
  - Cheat sheet http://shiny.rstudio.com/images/shiny-cheatsheet.pdf
  - Publish your app free http://www.shinyapps.io
  - Examples http://www.showmeshiny.com/
  - Tutorial by Dean Attali http://deanattali.com/blog/building-shiny-appstutorial/

