

Tech Skill Acquisition Master Class Series by STEMHub Foundation & AfricaHacks

Data Visualization with R studio & Tableau

Tech Skill Acquisition Masterclass Series by

STEMHub Foundation X AH! AfricaHacks

SERIES THREE

GOOGLE MEET meeting code: nqqxkxkrur

R

TOPIC

Data Visualization with Tableau and R studio

SPEAKERS:


Omotoyosi OGUNBANWO
Data Analyst (Finance) - Cisco


Mariam ADEYEMI
Financial Analyst & Founder of TechaVilly

May 30th, 2020
12PM EST || 5PM WAT

@stemhub_ @stemhubfoundation stemhubfoundation.com

@Africahacks Africahacks.com

Data Analytics & Visualization

with



Prepared by

Mariam Adeyemi

Financial Analyst &
Founder, TechaVilly



STEMHub
Foundation



AGENDA

- Introduction to Data Analytics
- Getting Started with the Basic Plots
- Introduction to Data Visualization in R
- Introduction to ggplot2 and how it works



Introduction to Business Analytics

Business Analytics revolves around problem solving. Providing solutions to business problems and answering the 4Ws and H questions.

In definition, Business analytics is a scientific process that transforms data into insights that are used for fact-based or data driven decision making.



Types of Business Analytics

- **Descriptive Analytics** which involves consolidation and summarization of data for further analysis. It includes techniques that explains what happened in the past
- **Predictive Analytics:** This is where the data received from descriptive analytics is used. The objective is to predict the future unforeseen events
- **Prescriptive Analytics** specifies the best course of action for a business activity in the form of output of a prescriptive model.



Introduction to Programming

R is a language and environment for statistical computing and graphics

R has many operators to perform different mathematical and logical operators.

A basic familiarity with R and its syntax will get you started easily



Installation of Software

Let's quickly go over the steps to install R

1. Go to R homepage and select CRAN
CRAN is an acronym for – Comprehensive R Archive Network. It's the collection of sites which carry R Distributions, Packages and documentation
2. Select the location which is nearest to you
3. Download and Install R depending in your Operating system



Installation of Software

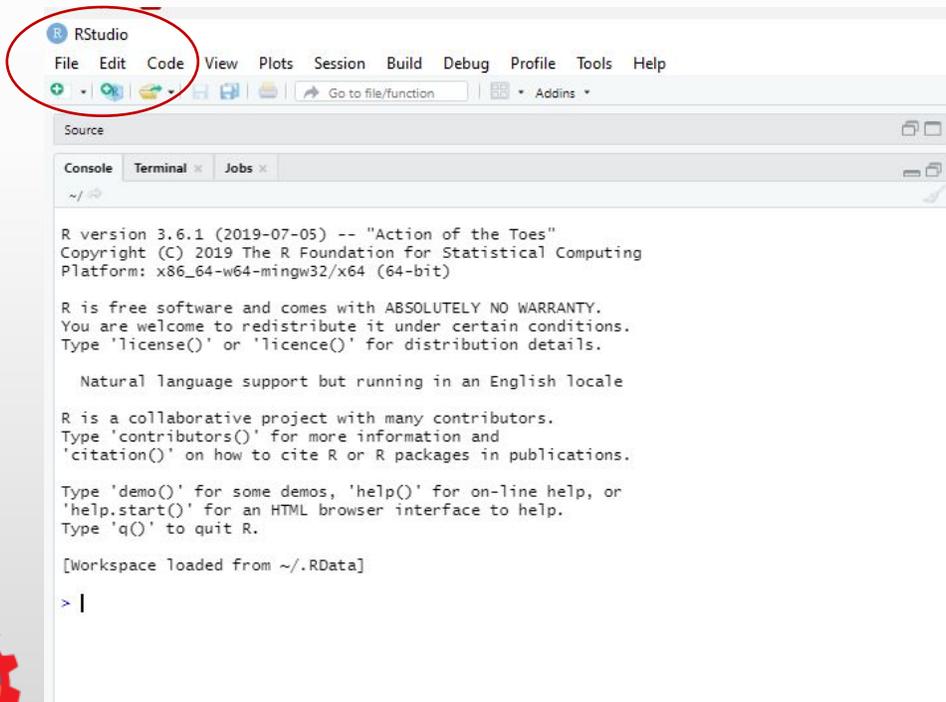
Once you are done installing, you can launch it from your Applications folder and type the following ;

```
> 1  
1  
> 'hello World'  
Hello world
```



Loading Data into

Datasets can either be built-in or can be loaded from external sources.



The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The console window displays the following text:

```
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]
> |
```

You can also use
any of the in-built
datasets to
practice at your
free time



Loading Data into



You can view your data frame by typing the code below:

```
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

```
[Workspace loaded from ~/.RData]
```

```
> data("airquality")  
> head(airquality)  
  Ozone Solar.R Wind Temp Month Day  
1    41    190  7.4   67     5   1  
2    36    118  8.0   72     5   2  
3    12    149 12.6   74     5   3  
4    18    313 11.5   62     5   4  
5    NA     NA 14.3   56     5   5  
6    28     NA 14.9   66     5   6  
> |
```

Use this code to load the dataset

Use this code to view the data frame

And this is what the data looks like



Introduction to Programming

There are four basic types operators in R

1. Arithmetic Operators
2. Relational Operators
3. Logical Operators
4. Assignment Operators



Arithmetic Operators

This is used for mathematical operators like the additions and subtractions.

Operators	Descriptions
+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Exponent
%%	Modulus
%/%	Integer Division



Arithmetic Operators Example

$x = 10, y = 3$

Operators	Code	Output
+	$x+y$	13
-	$x-y$	7
*	$x*y$	30
/	x/y	3.333
\wedge	$x^\wedge y$	1000
%%	$x\%y$	1 (x into y remainder)
%/%	$X\%/y$	3(x into y and how many times its repeated)



Relational Operators

These are used to compare two values that return a logical output

Operators	Descriptions
<	Less than
>	Greater than
<=	Less than or Equal to
>=	Greater than or Equal to
==	Equal to
!=	Not Equal to



Relational Operators Example

$x = 10, y = 3$

Operators	Code	Output
<	$2 < 3$	TRUE
>	$2 > 3$	FALSE
<=	$2 <= 2$	TRUE
>=	$3 >= 3$	TRUE
==	$x = 3^4, x == 81$	TRUE
!=	$2 != 3$	TRUE



Logical Operators

These are used to perform Boolean operators such as “AND” and “OR”

Operators	Descriptions
!	Logical Not
&	Element-wise logical AND
&&	Logical AND
	Element-wise logical OR
	Logical OR



Logical Operators Example

For instance, we can name a variable FALSE or TRUE

E.g: `aVariable <- FALSE`. If we run this in the R software, it will return an output that says “Logical”

`x = 1, y = 2`

Operators	Code	Output
!	<code>x!=10</code>	TRUE
&	<code>X==1, y==2</code>	TRUE
&&	<code>X<1 && y>4</code>	FALSE
	<code>X==1 y==3</code>	TRUE
	<code>X=2 y==2</code>	TRUE



Assignment Operators

These are used to assign values to variables.

It is advisable to use the left assignment with an arrow

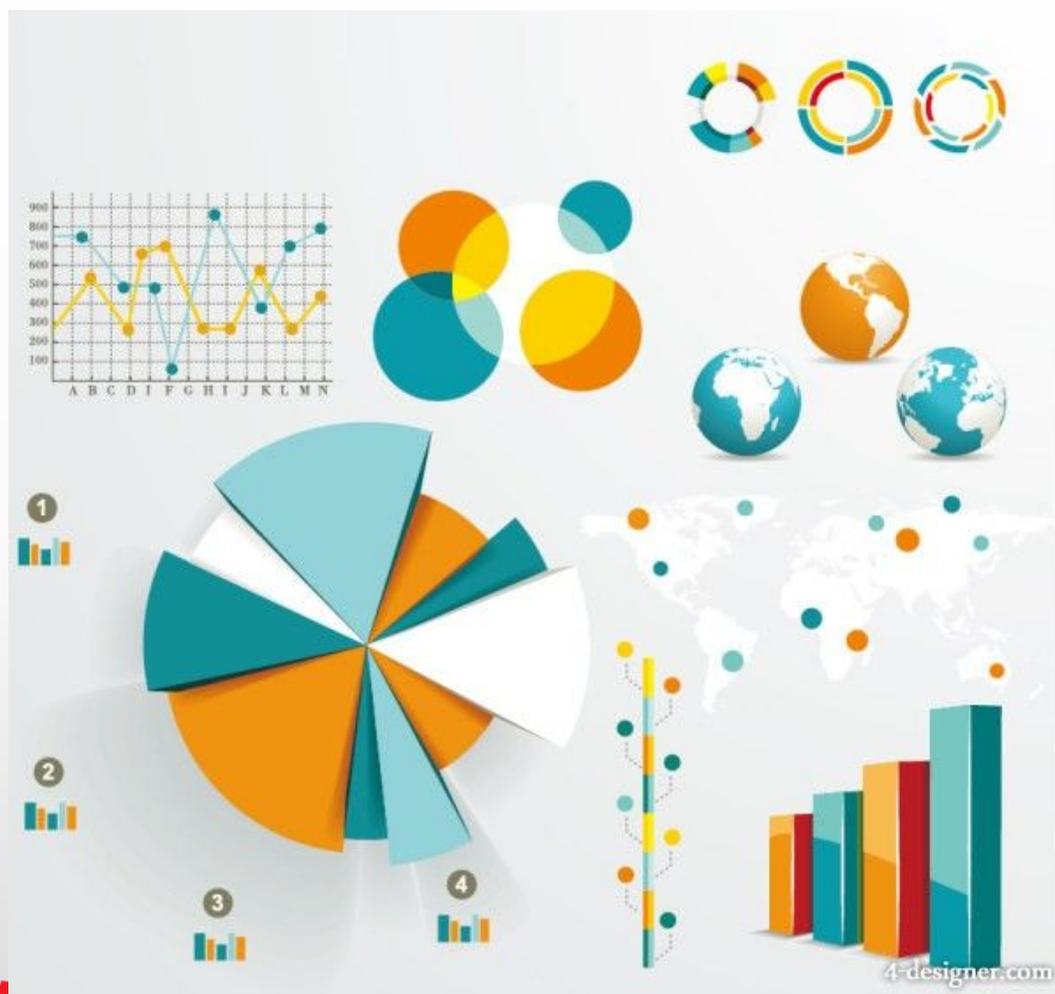
Operators	Descriptions
<-	Left arrow

See example below;

```
Age <- 18
```



Data Visualization in R



It is “not” possible to tell stories without visualizations with the volume of data available in the world today.

One of the greatest advantages of data visualization is that it helps tremendously to explain complex datasets.

R software is one of the tools that does this well.



Basic Visualization tools in R

I will be using 4 basic Visualization types in this session

- Histogram
- Box plot
- Scatter plot
- Line Graph



Basic Visualization

The Histogram

The histogram is basically a plot that breaks the data into bins (or breaks) and shows frequency distribution of these bins.

The Box plot

Box plot shows 5 statistically significant numbers – the 25th percentile, the 75th percentile and the maximum. It is useful for visualizing the spread of the data and deriving inference accordingly.



Basic Visualization

The Scatter Plot

The scatter plots help in visualizing data easily and for simple data inspection.

The Line Chart

Line charts are used to analyze a trend over a period of time. You can also use line chart when comparing relative changes in quantities across some variables.



Conclusion

We've seen how easy and simple Data Visualization is, in R, whether you are using the in-built data, or you are using your own data.

Obviously, visualization capabilities in R can not be underestimated!



The Codes

Histogram

```
hist(airquality$Solar.R,main = 'Solar Radiation values in air',xlab = 'Solar rad',col = 'red')
```

Vertical bar plot

```
# Vertical bar plot > barplot(airquality$Ozone,main = 'Ozone Concentration in air',xlab = 'Ozone levels',col='red',horiz = FALSE)
```

Horizontal bar plot

```
# Horizontal bar plot > barplot(airquality$Ozone,main = 'Ozone Concentration in air',xlab = 'Ozone levels',col = 'green',horiz = TRUE)
```



The Codes

colored scatter plot

```
ggplot(data = mtcars, mapping =  
aes(x=wt, y=mpg, color=as.factor(cyl)))+geom_point()
```

Line graph

```
ggplot(data = mtcars, mapping = aes(x=wt, y=mpg))+geom_abline()
```

Map

```
map(database = "state")
```



The Codes

To load the data

```
data(airquality)
```

To view the data frame

```
head(airquality)
```



THANK YOU



STEMHub
Foundation



TABLEAU FOR DATA ANALYSIS

BY OMOTOYOSI OGUNBANWO (TechAvilly)



STEMHub
Foundation



Business Intelligence

- Tableau is a business intelligence tool used for advanced data analysis and visualization. Other business intelligence tools include PowerBI, Oracle, microstrategy...etc
- Prior to tableau, most analysts build dashboards on excel and transfer to power point for presentations. However Tableau can provide both in one
- Tableau can help provide great insights into customer data that you normally wouldn't be able to understand
- As a business owner, you are typically faced with multiple questions like how to improve customer satisfaction and increase revenue. You can't answer these questions without the relevant data to help you understand
- At the end of this class, you will be able to create a simple dashboard for presentation.



Why Tableau



Unlike Excel spreadsheet where you have to manually enter your data, Tableau can be connected to multiple data sources like Oracle, SQL, Access and even excel.



Dashboards are very easy to build on tableau. You can create a new insight to your data in less than 10 seconds with tableau.



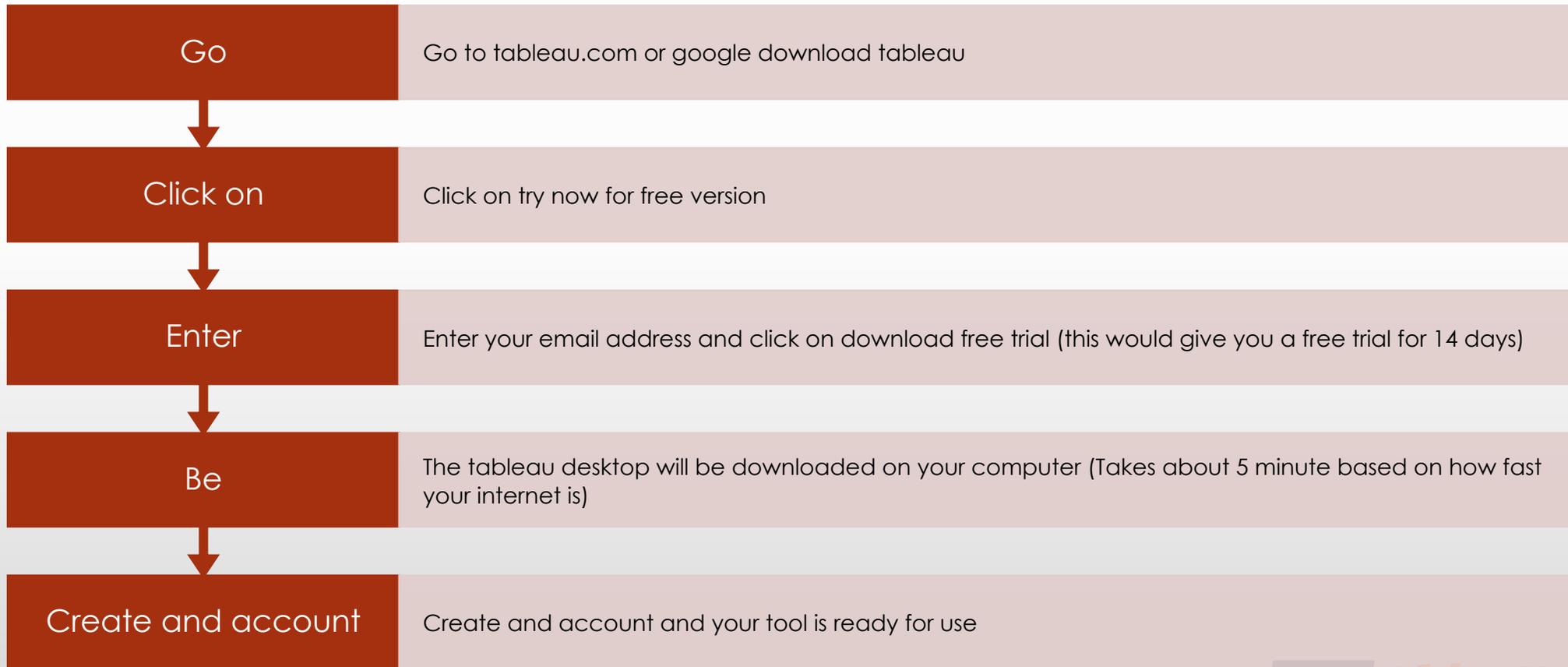
Tableau can help you with forecasting and predicting



However, the best analyst is the one who can combine both excel and tableau together.



Downloading Tableau



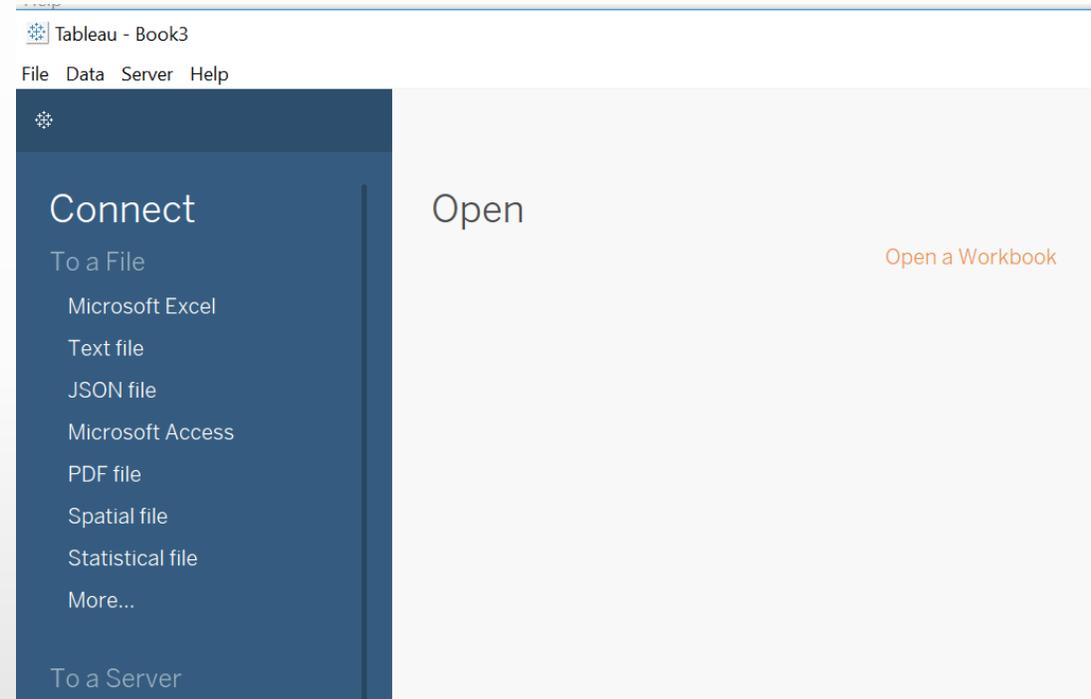
SAMPLE DATA SET

- We are going to use some sample data sets for our class today. We want this interactive so everyone can participate.
- Data set 1
- <https://community.tableau.com/docs/DOC-10198>
- Data set 2
- <https://data.gov.sg/dataset/resident-population-by-ethnicity-gender-and-age-group>



Using Tableau

- Once you open your newly downloaded tableau, to your left, you will see the servers and sheets you can connect to.
- Since we will be working on excel today, we are going to connect our tableau to excel to build our new dashboard



Using Tableau cont'd



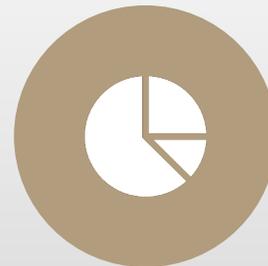
Now that we have successfully connected our excel sheet to tableau, lets try and understand our new dashboard before building our visualization



On the tableau dashboard, we have the Dimensions Vs Measures. Also the show me tab for our various visualizations



Dimensions in tableau are our independent variables. Also our descriptive variables



Measures are the quantitative and statistical count of the dimensions. Measures contain sum, average, mean, standard deviation, etc.

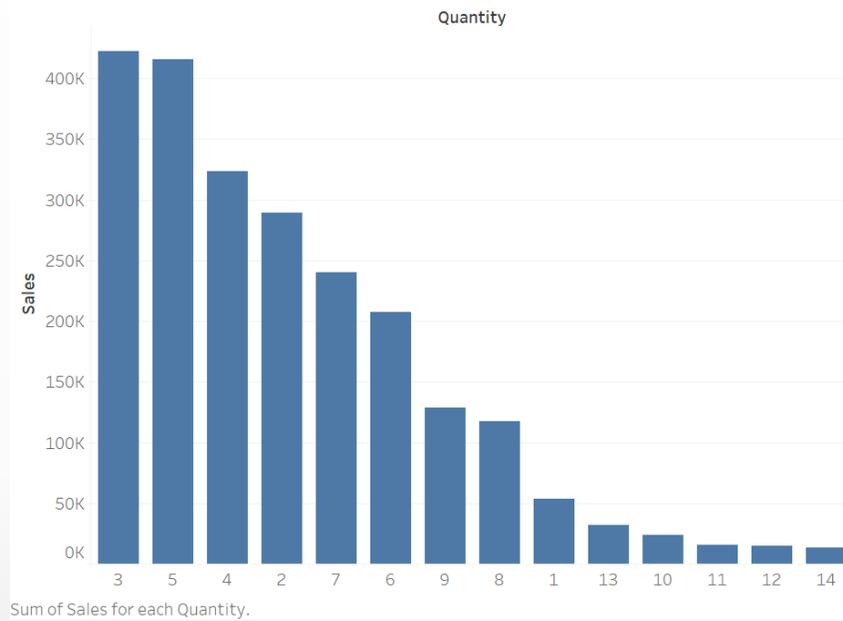


Visualization

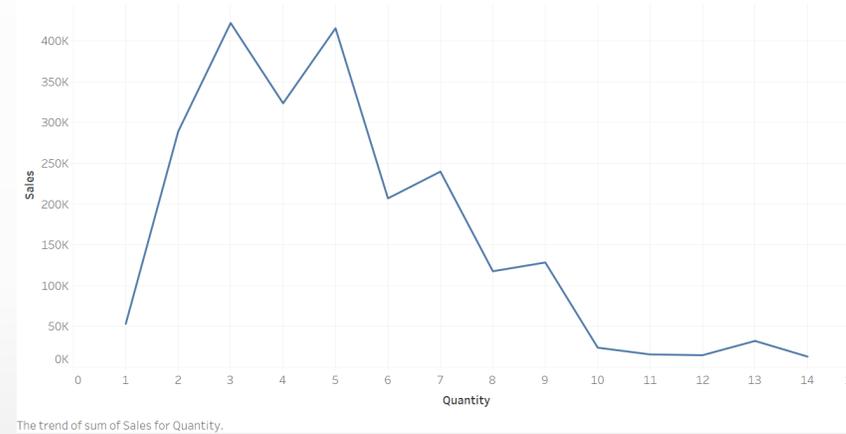
- Anything that can be aggregated is a measure. i.e our region and quantity example shows us how dimension and measures work
- Tableau is smart enough to identify dimensions and measures. However if something falls as a dimension when it should be a measure, you can always drag it.
- Blue is dimension and green is measures. Keep this in mind as it helps when analyzing large data sets
- Now we are going to show the difference between a continuous graph and a discrete graph. In statistics, discrete variables are finite while continuous variables are infinite



Sheet 2

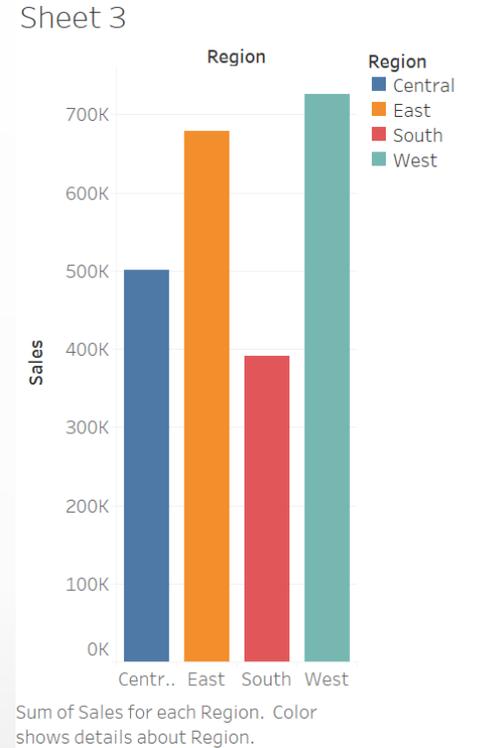
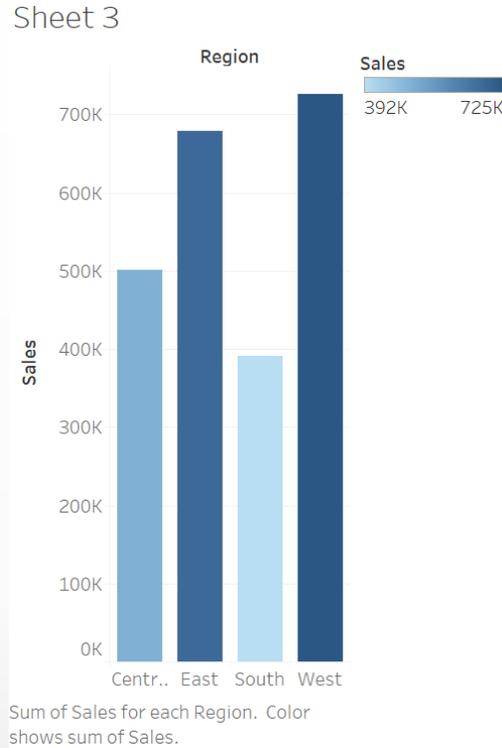


Sheet 2



Discrete Vs Continuous



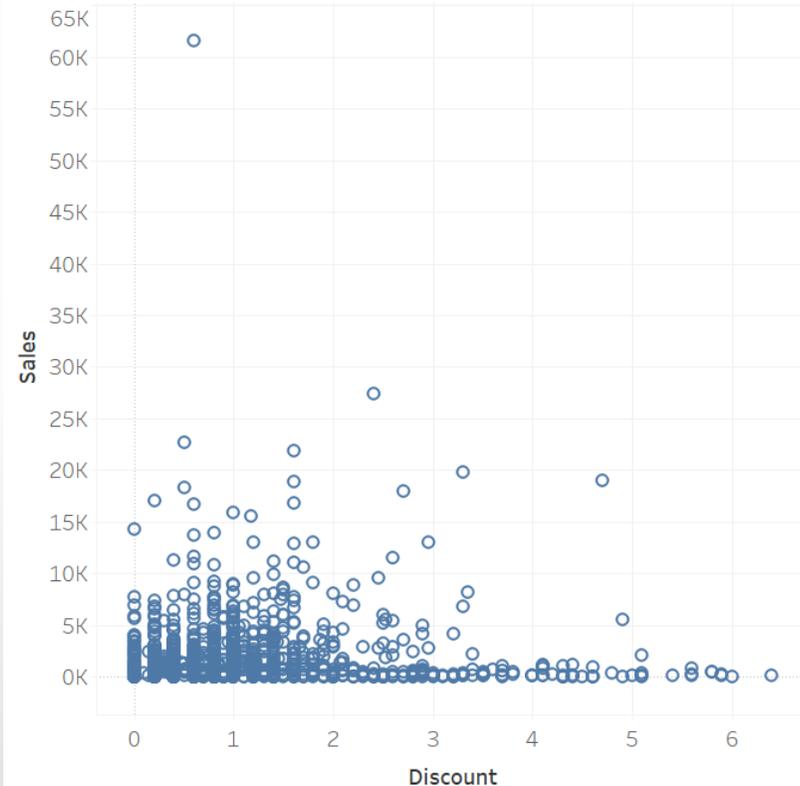


GRAPHS



SCATTER PLOT

Sheet 4

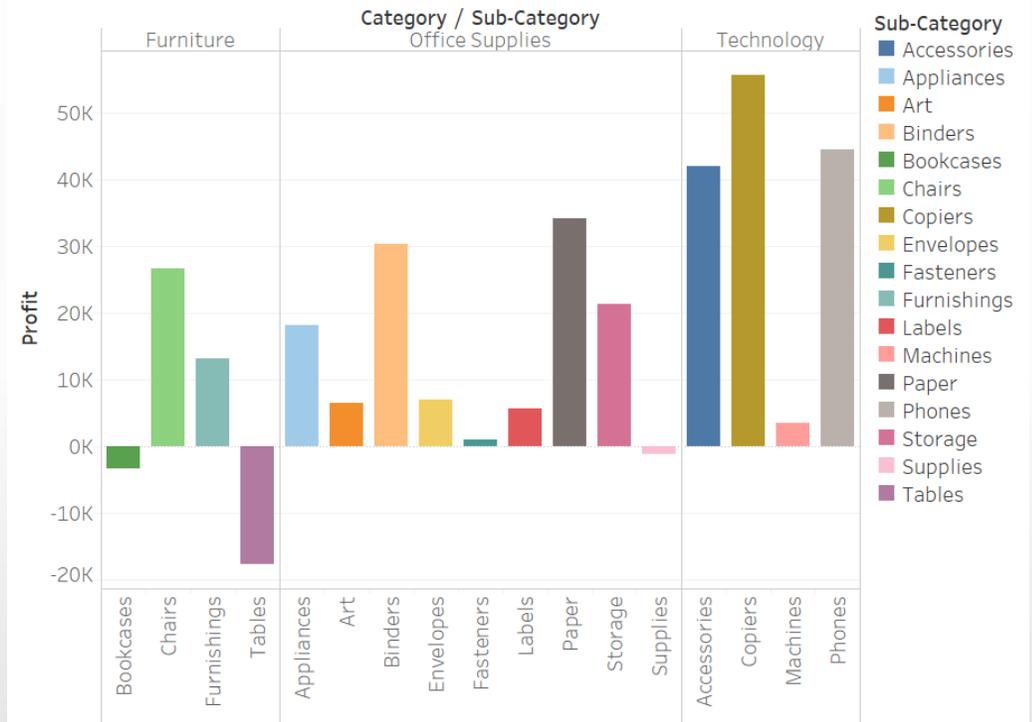


Sum of Discount vs. sum of Sales. Details are shown for Product ID.

- A scatter plot is made up of two measures
- One measure goes to the Y axis rows (sales) and the other to the X axis columns (discount)
- From this plot, we can see that products with high sales have lower discounts.



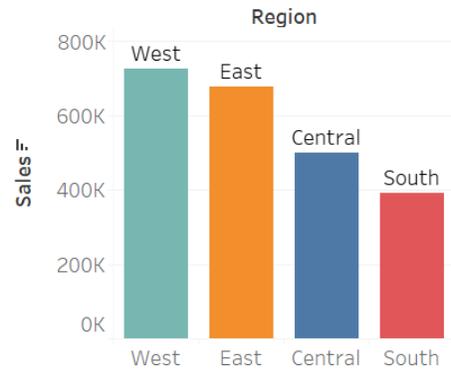
Sheet 5



Sum of Profit for each Sub-Category broken down by Category. Color shows details about Sub-Category. Details are shown for Sub-Category.

PROFIT BY SUB CATEGORY

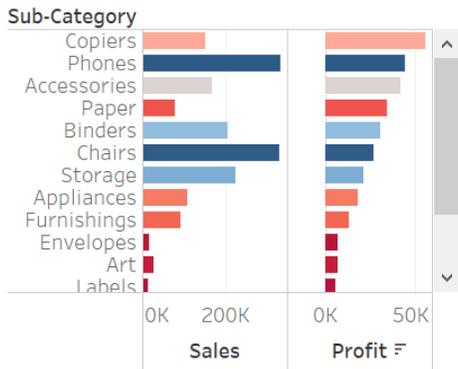
Sheet 3



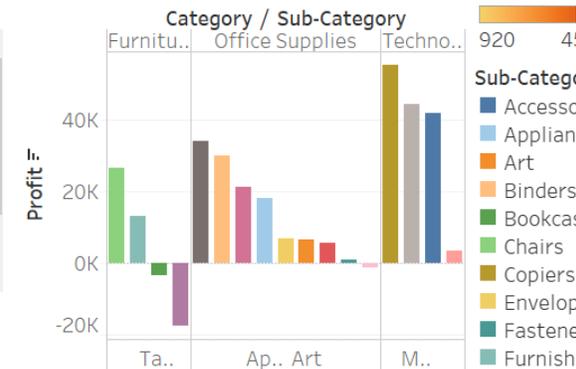
Sheet 6



Sheet 7



Sheet 5



Creating a Dashboard



TABLEAU PRACTICE

- I want you all to practice this data set and replicate the below visualization
- The first person to replicate this will be awarded TechAville Data Champion of the week.
- TechAville Data Champion of the week gets featured on our website <http://www.techavilly.com/>
- The data champion of the week gets a \$50 cash prize plus access to our mentors.
- We will help you through transitioning to STEM. And if you are already in STEM, we provide additional guidance on how to improve productivity and become a high achiever.



Tableau Practice

- To participate, kindly follow the below link and watch the Tableau video posted <https://www.youtube.com/watch?v=ap0QC8HbeWA&t=61s>
- After watching the video, follow the below link to download the required data set for the visualization. **<https://www.techavilly.com/tableau-dataset-download>**
- Then replicate the HD visualization images attached below in this presentation.
- The first person to complete and email to toyosiogunbanwo@gmail.com and aderonke12@yahoo.com will be the winner for the week.
- Please note that you must send your final answer to both emails above at the same time.
- You must replicate all the images the exact way below. Your email would have 5 attachments. 4 replicated screen snipe like the below images and a tableau file containing your complete work



Device Preview

Size
min 420x560 - max 650x8...

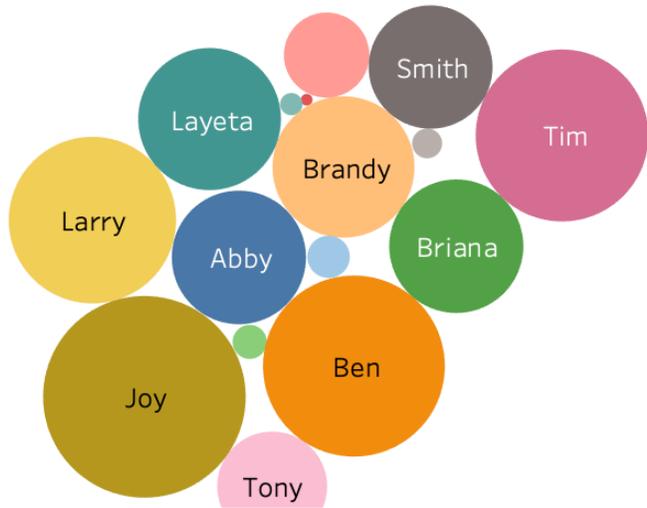
Sheets

- Sheet 1
- Sheet 2
- Sheet 3
- Sheet 4

Objects

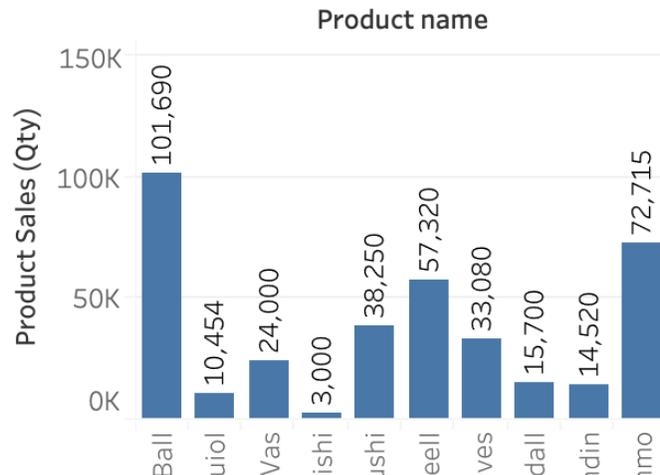
- Horizontal
- Vertical
- Text
- Image
- Web Page
- Blank
- Button

Tiled Floating

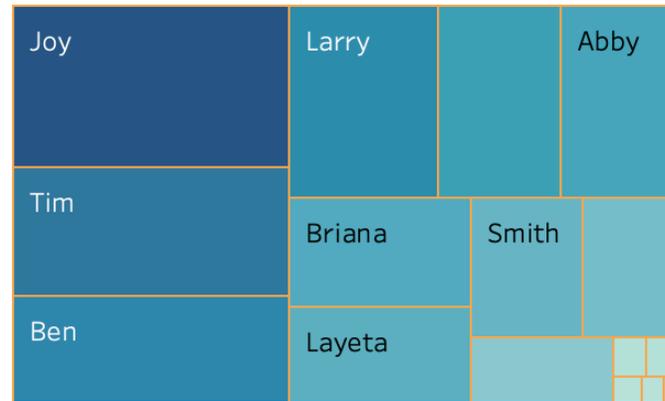


Product na..	State			
	Califonia	Colorado	Florida	Texas
BinBall	9,370	650		91,67
Nuiol	2,300	900		7,25
TadVas				24,00
Tamishi				3,00
Tamushi				38,25
Tuneell				57,32
Unves				33,08
Vebdall				13,00
vondin			1,750	12,77
Yunmo				72,71

Sheet 1



Sheet 4



Dimensions

- Date
- Product ID
- Product name
- Promotion
- Sales Man
- State, County
 - State
 - County
- Measure Names

Measures

- % Grossprofit
- % Sales
- Cost per product
- Monthly Baseline
- Product Sales (Qty)
- Product Sales \$
- Total Product Cost
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Pages

Columns Product name

Rows SUM(Product Sales (..

Filters

Marks

Automatic

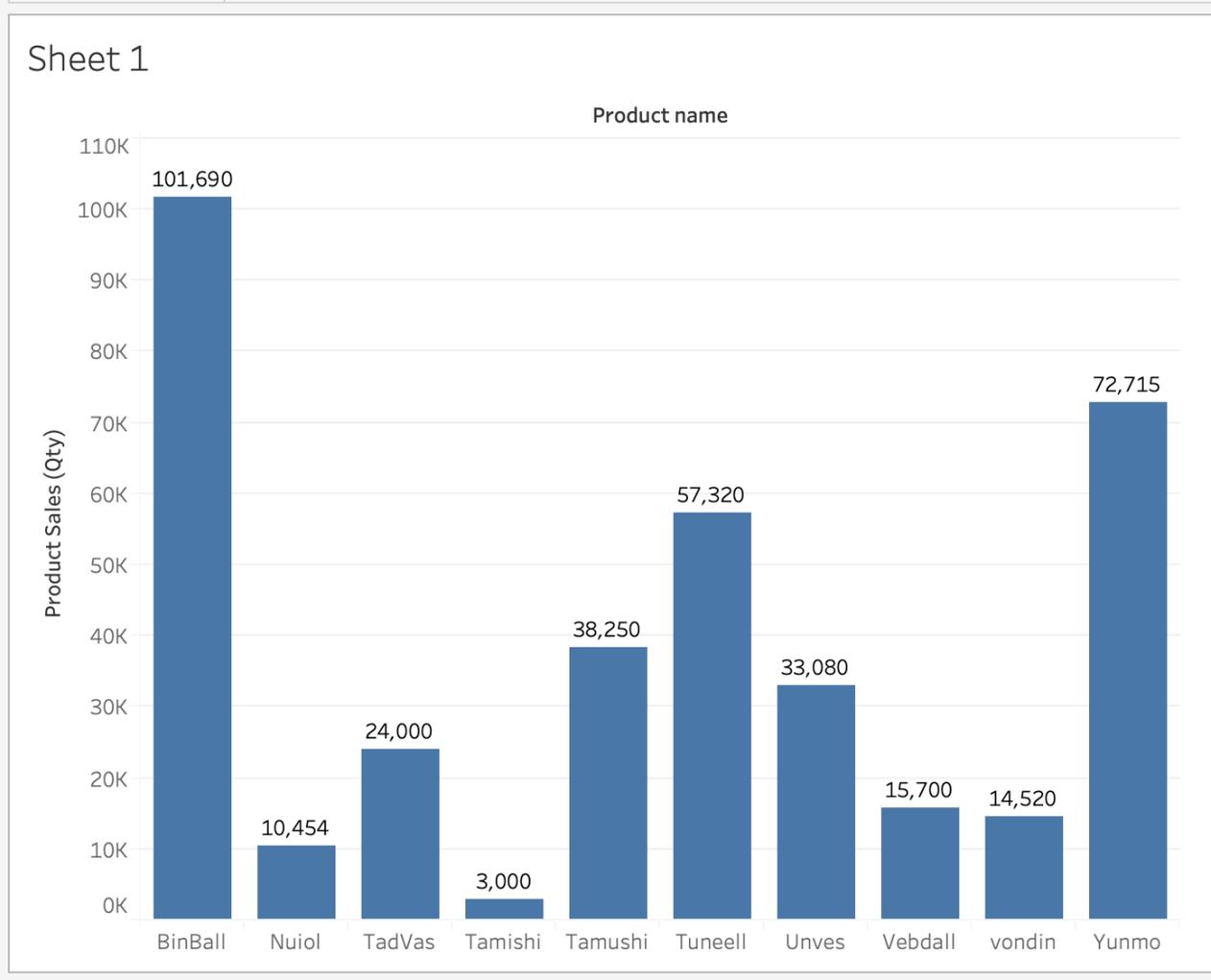
Color Size Label

Detail Tooltip

SUM(Product ..

Columns Product name

Rows SUM(Product Sales (..

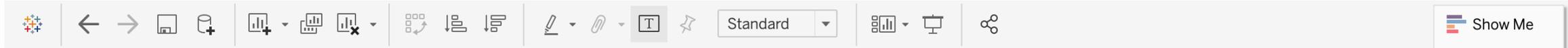


For horizontal bars try

0 or more Dimensions

1 or more Measures





Data | Analytics

Sales by Salesman Doug...

Dimensions

- Date
- Product ID
- Product name
- Promotion
- Sales Man
- State, County
 - State
 - County
- Measure Names

Measures

- % Grossprofit
- % Sales
- Cost per product
- Monthly Baseline
- Product Sales (Qty)
- Product Sales \$
- Total Product Cost
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Columns

Rows

Filters

Marks

Circle

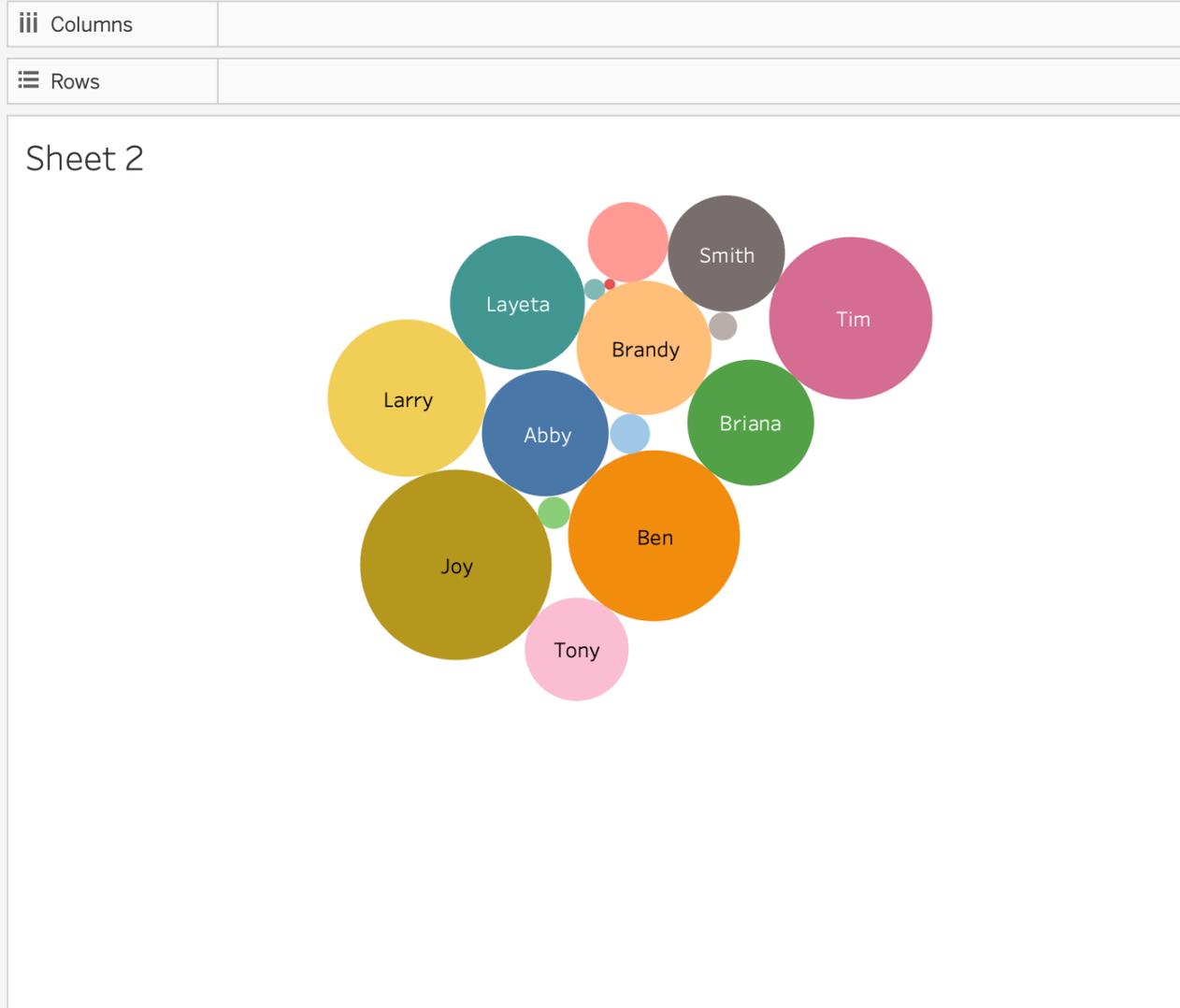
Color | Size | Label

Detail | Tooltip

SUM(Product ..)

Sales Man

Sales Man



Show Me

For **horizontal bars** try

0 or more **Dimensions**

1 or more **Measures**





Show Me

Data Analytics ▾
Sales by Salesman Doug...

Dimensions ▾

- Date
- Product ID
- Product name
- Promotion
- Sales Man
- State, County ▾
 - State
 - County

Measure Names

Measures

- % Grossprofit
- % Sales
- Cost per product
- Monthly Baseline
- Product Sales (Qty)
- Product Sales \$
- Total Product Cost
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Pages

Columns: State

Rows: Product name

Filters

Marks

Square

Color Size Label

Detail Tooltip

SUM(Product ..)

SUM(Product ..)

Sheet 3

Product na..	State				
	California	Colorado	Florida	Texas	Washing..
BinBall	9,370	650		91,670	
Nuiol	2,300	900		7,254	
TadVas				24,000	
Tamishi				3,000	
Tamushi				38,250	
Tuneell				57,320	
Unves				33,080	
Vebdall				13,000	2,700
vondin			1,750	12,770	
Yunmo				72,715	

For **symbol maps** try

- 1 geo Dimension
- 0 or more Dimensions
- 0 to 2 Measures

May use spatial measure in place of geo dimension

References



Simplilearn tableau training by Simplilearn



Tableau in 2 minutes by penguin analytics



Tableau full course by intellipart



Social Media Handles

For More Details, please follow:

STEMHub Foundation:

IG: @stemhubfoundation

Twitter: @stemhub_

LinkedIn: STEMHub Foundation

Facebook: STEMHub Foundation

Website: stemhubfoundation.com

For More Details, please follow:

AfricaHacks

IG & Twitter: @Africa_hacks

LinkedIn: AfricaHacks

Facebook: AfricaHacks

Website: africahacks.com



THANK YOU



STEMHub
Foundation

