



TechTalent Academy Safeguarding Policy

*“Protecting an adult’s right to live in **safety, free from abuse and neglect**. It is about people and organisations working together to **prevent and stop both the risks and experience of abuse or neglect**, while at the same time making sure that the **adult’s wellbeing is promoted** including, where appropriate, having regard to their views, wishes, feelings and beliefs in deciding on any action. This must recognise that adults sometimes have complex interpersonal relationships and may be ambivalent, unclear or unrealistic about their personal circumstances.”*

If you have a safeguarding concern, please raise this with your tutor or via the safeguarding link on our website:

<https://www.techtalent.co.uk/safeguarding-statement>

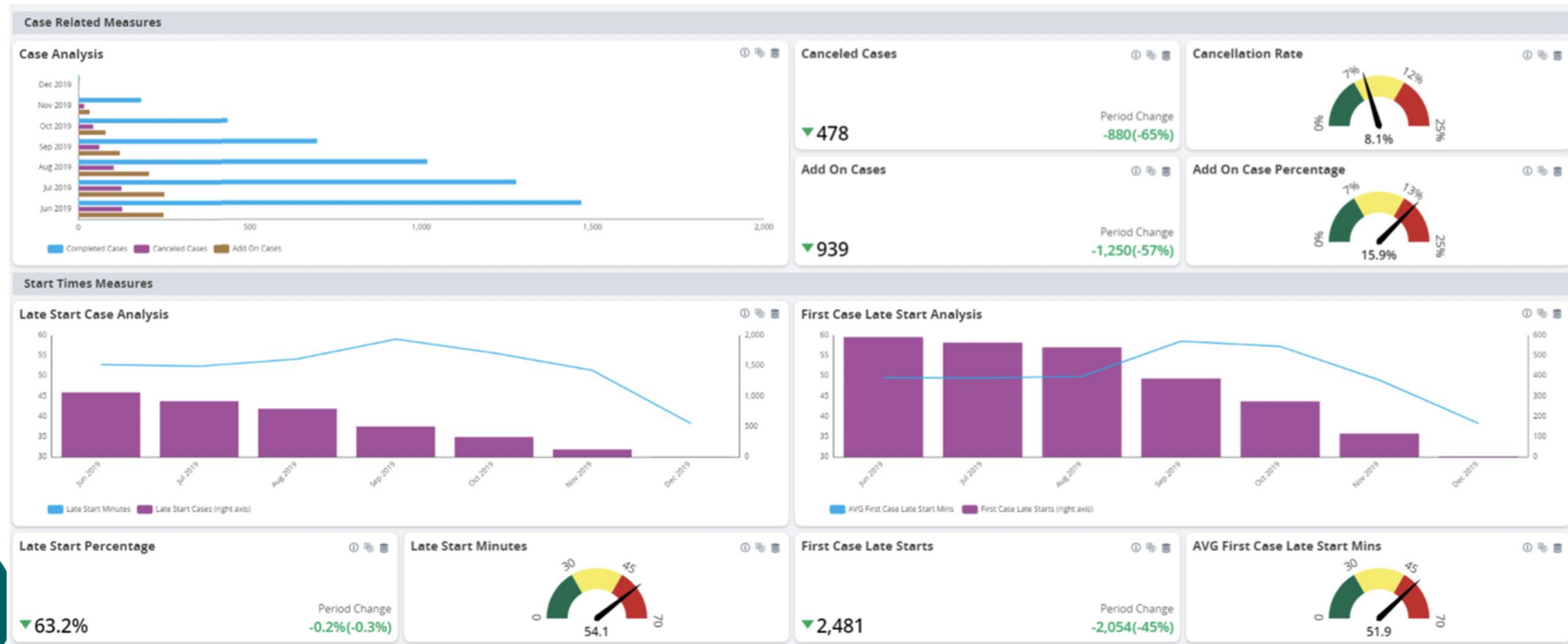
TechTalent’s safeguarding lead is: **Max Ruddock**





Starter Activity.

What is the purpose of a dashboard?





TechTalent Academy

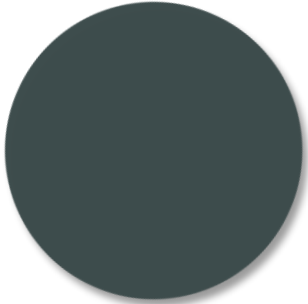


Data Science Course

Data Visualisation



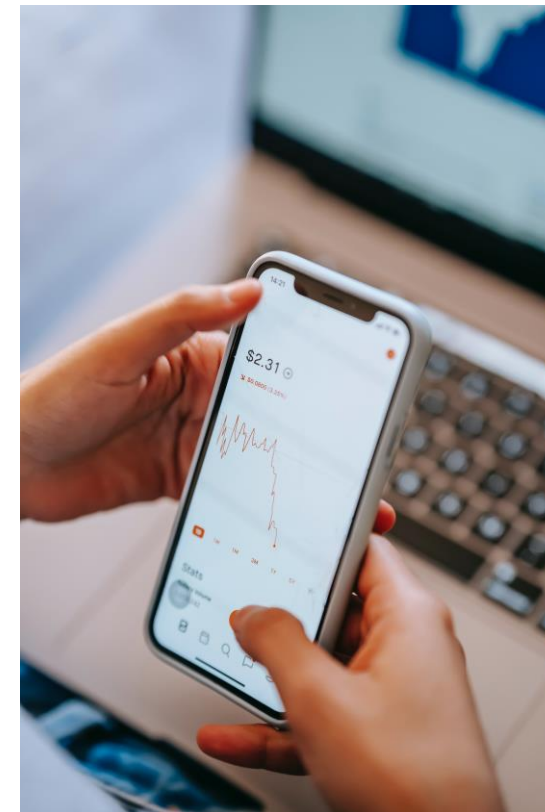


Lesson Objectives.

- 
- How to choose the right visualisation?
 - What is Tableau?
 - How to make a dashboard ?
- 
- 

Data Visualisation Theory.

- Data visualisation studies is the field of study that focuses on the graphic representation of data.
- In creating data visualisations, we intend to communicate information to an audience or user in a digestible way.



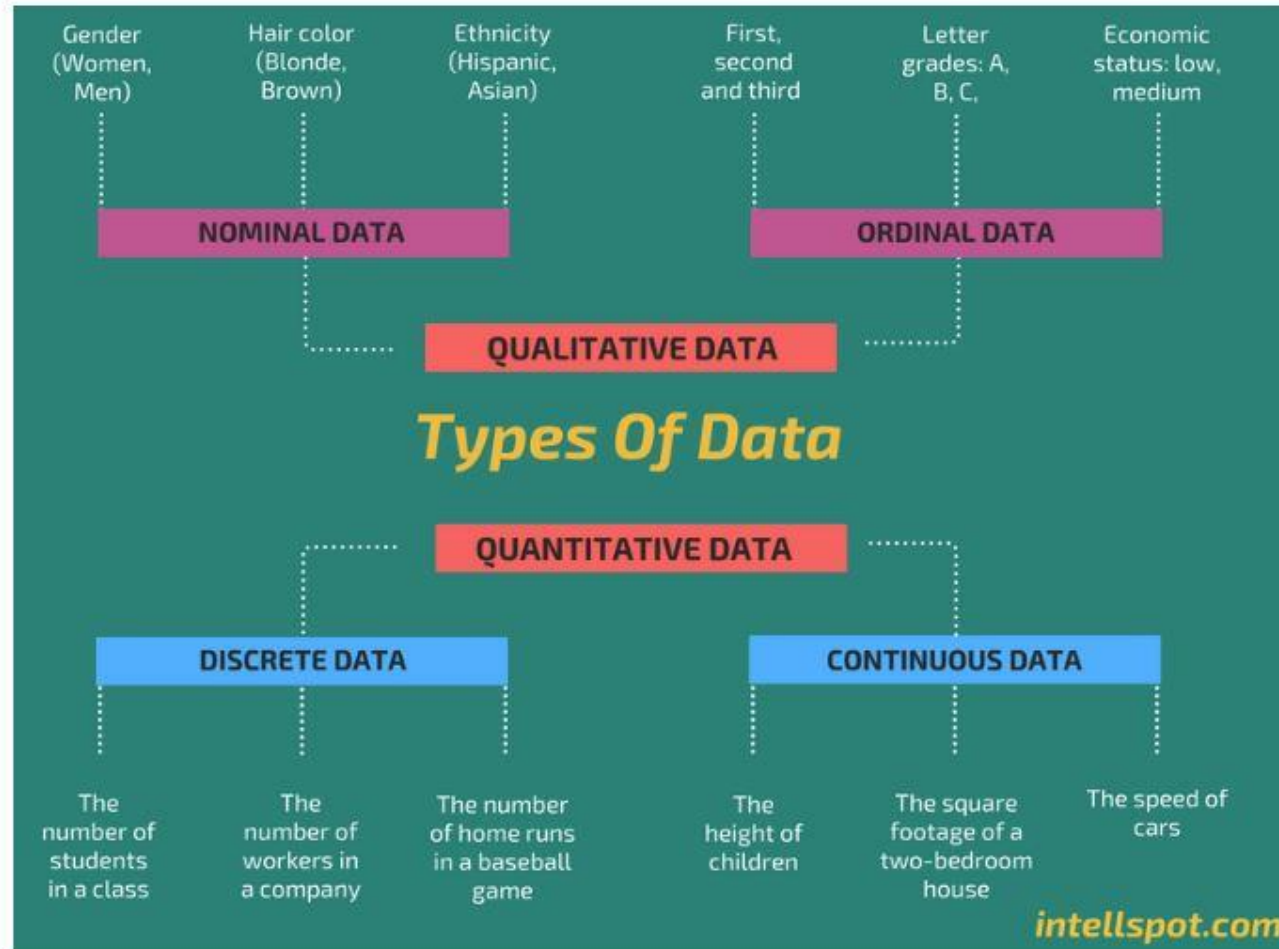
Choosing The Right Visualisation.

Choosing the right type of chart/graph will depend mostly on the types of variables you are analysing from your dataset.



Types of variables.

Choosing the right data visualisation plot depends on the variable type :



Types of variables.

Categorical/Qualitative i.e. groups : gender/sex, Score (excellent, good, pass, fail)

- nominal**: a name or label, no order (sex/gender)

- ordinal**: order (Score :excellent, good, pass, fail)

Numeric/Quantitative: real amount of data that you can make mathematical operations on

- continuous**: take infinite number of real values (age, height, weight, distance)

- discrete** : take finite number of real values (score from 0 to 10, number of individuals in a household)

Independent vs dependant variables:

- Dependent variable: the one being measured during an experiment/study, their values depends on other variables (typically on Y axis)

- Independent variable will have an effect on the dependant variable, stand alone variable, not being influenced by other variables (typically on X axis)

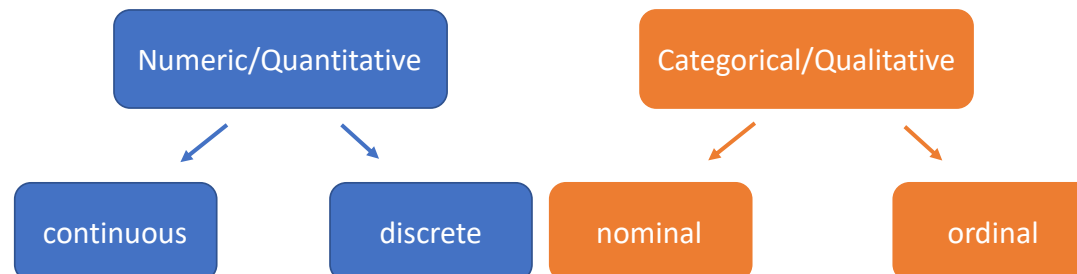
Task.

In which category would you put the following variables:

| Examples | Variable type? |
|--------------------------------|----------------|
| Hair colour | |
| # of employees in a company | |
| Ethnicity | |
| Weight | |
| Grades (Excellent, good, pass) | |
| Speed of a motorcycle | |

5:00

*5 Minute
Countdown Timer*



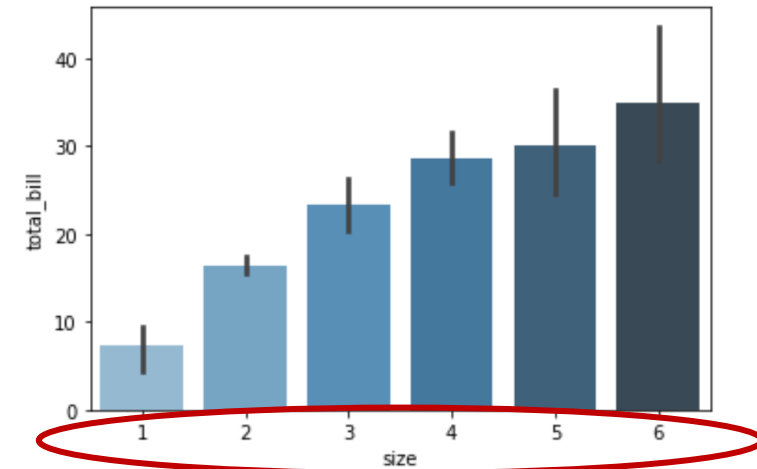
Choosing The Right Visualisation.

Bar plots: ideal for comparing discrete categorical variables (sex, smoker, day, time, size)

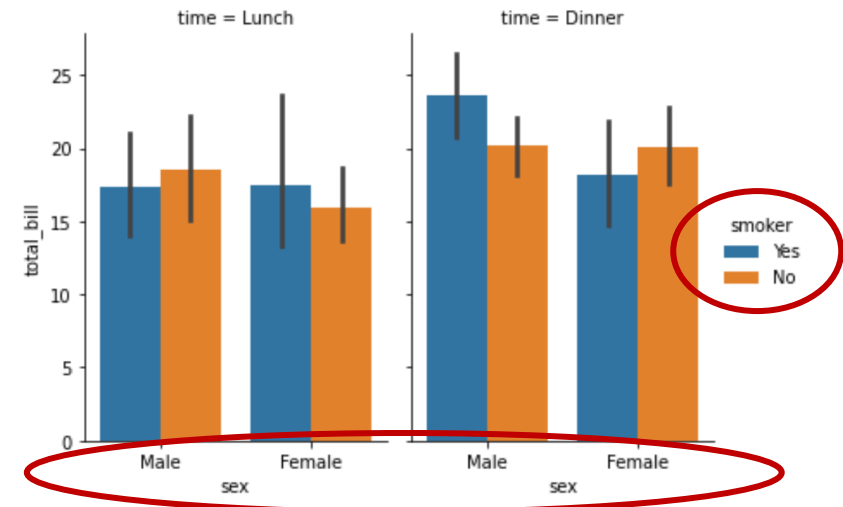
| | total_bill | tip | sex | smoker | day | time | size |
|-----|------------|------|--------|--------|------|--------|------|
| 0 | 16.99 | 1.01 | Female | No | Sun | Dinner | 2 |
| 1 | 10.34 | 1.66 | Male | No | Sun | Dinner | 3 |
| 2 | 21.01 | 3.50 | Male | No | Sun | Dinner | 3 |
| 3 | 23.68 | 3.31 | Male | No | Sun | Dinner | 2 |
| 4 | 24.59 | 3.61 | Female | No | Sun | Dinner | 4 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 239 | 29.03 | 5.92 | Male | No | Sat | Dinner | 3 |
| 240 | 27.18 | 2.00 | Female | Yes | Sat | Dinner | 2 |
| 241 | 22.67 | 2.00 | Male | Yes | Sat | Dinner | 2 |
| 242 | 17.82 | 1.75 | Male | No | Sat | Dinner | 2 |
| 243 | 18.78 | 3.00 | Female | No | Thur | Dinner | 2 |

244 rows × 7 columns

Individual Bar plot



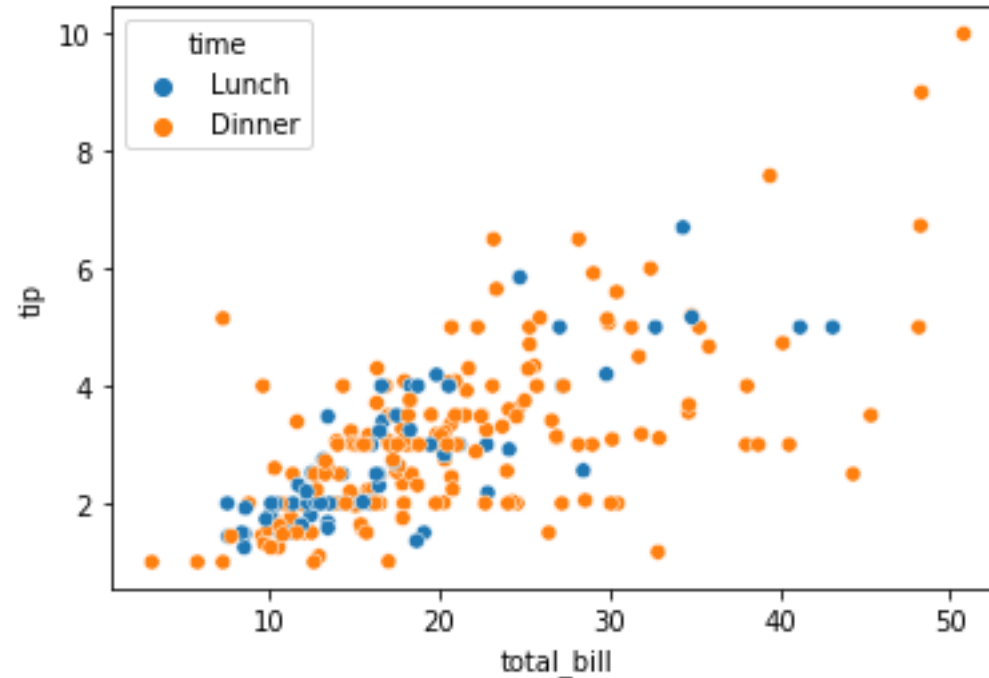
Grouped Bar plot



Choosing The Right Visualisation.

Scatter plot:

investigate the relationship between 2 variables X and Y. Each value from the dataset represents a dot on the graphic.

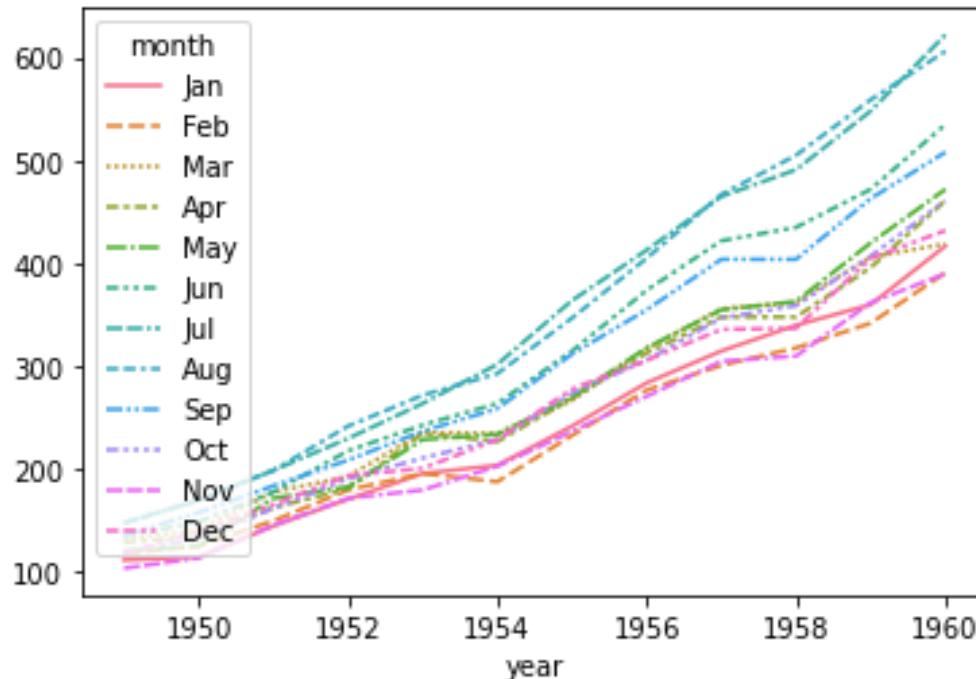


Choosing The Right Visualisation.

Flight passengers over months and years

Line chart: ideal for following how continuous variables changed over time, or finding trends over time

| month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| year | | | | | | | | | | | | |
| 1949 | 112 | 118 | 132 | 129 | 121 | 135 | 148 | 148 | 136 | 119 | 104 | 118 |
| 1950 | 115 | 126 | 141 | 135 | 125 | 149 | 170 | 170 | 158 | 133 | 114 | 140 |
| 1951 | 145 | 150 | 178 | 163 | 172 | 178 | 199 | 199 | 184 | 162 | 146 | 166 |
| 1952 | 171 | 180 | 193 | 181 | 183 | 218 | 230 | 242 | 209 | 191 | 172 | 194 |
| 1953 | 196 | 196 | 236 | 235 | 229 | 243 | 264 | 272 | 237 | 211 | 180 | 201 |

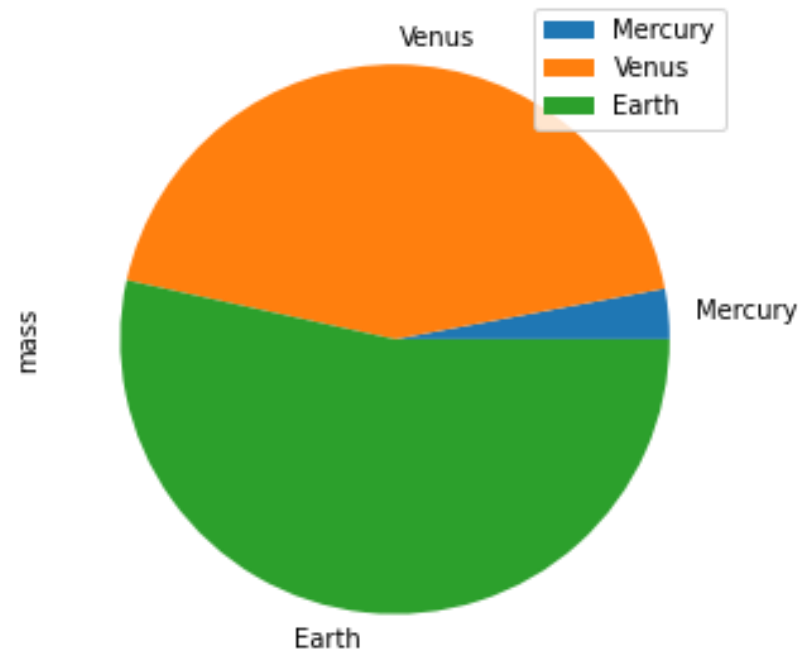


Time series

Choosing The Right Visualisation.

Pie chart: ideal for representing different proportions of variables

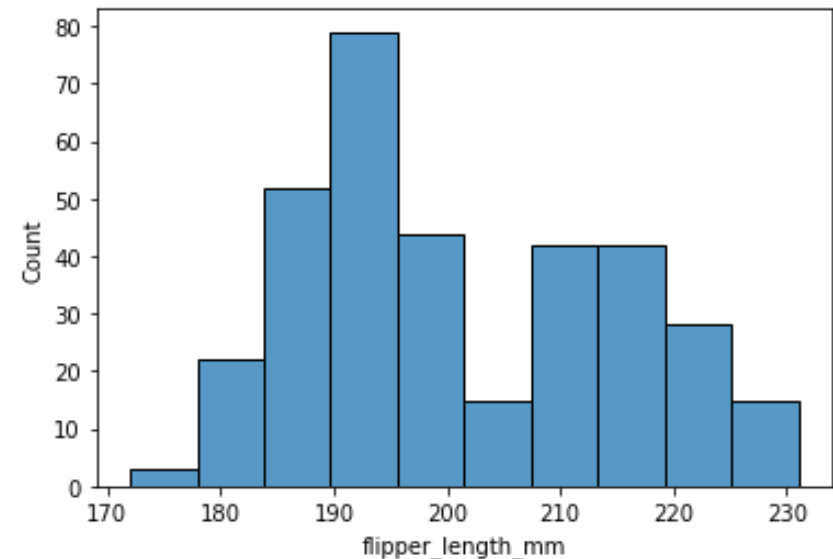
| | mass | radius |
|---------|------|--------|
| Mercury | 0.33 | 2439.7 |
| Venus | 4.87 | 6051.8 |
| Earth | 5.97 | 6378.1 |



Choosing The Right Visualisation.

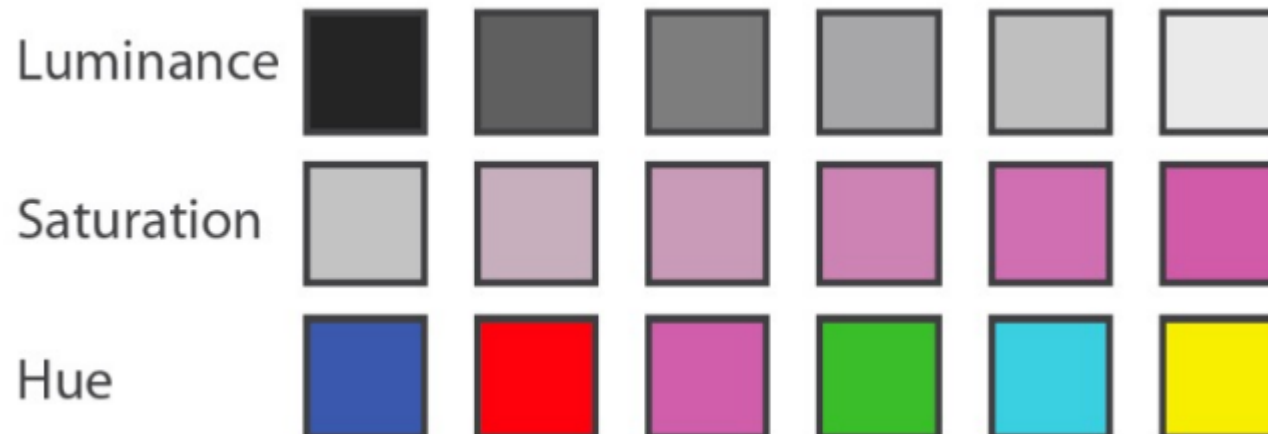
Histogram: ideal to observe the distribution of a variable

| | species | island | bill_length_mm | bill_depth_mm | flipper_length_mm | body_mass_g | sex |
|-----|---------|-----------|----------------|---------------|-------------------|-------------|--------|
| 0 | Adelie | Torgersen | 39.1 | 18.7 | 181.0 | 3750.0 | Male |
| 1 | Adelie | Torgersen | 39.5 | 17.4 | 186.0 | 3800.0 | Female |
| 2 | Adelie | Torgersen | 40.3 | 18.0 | 195.0 | 3250.0 | Female |
| 3 | Adelie | Torgersen | NaN | NaN | NaN | NaN | NaN |
| 4 | Adelie | Torgersen | 36.7 | 19.3 | 193.0 | 3450.0 | Female |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 339 | Gentoo | Biscoe | NaN | NaN | NaN | NaN | NaN |
| 340 | Gentoo | Biscoe | 46.8 | 14.3 | 215.0 | 4850.0 | Female |
| 341 | Gentoo | Biscoe | 50.4 | 15.7 | 222.0 | 5750.0 | Male |
| 342 | Gentoo | Biscoe | 45.2 | 14.8 | 212.0 | 5200.0 | Female |
| 343 | Gentoo | Biscoe | 49.9 | 16.1 | 213.0 | 5400.0 | Male |



Decomposing Colour.

- Hue can represent categorical information
- Luminance and saturation can show ordered information
- Should only be for a finite number of bins



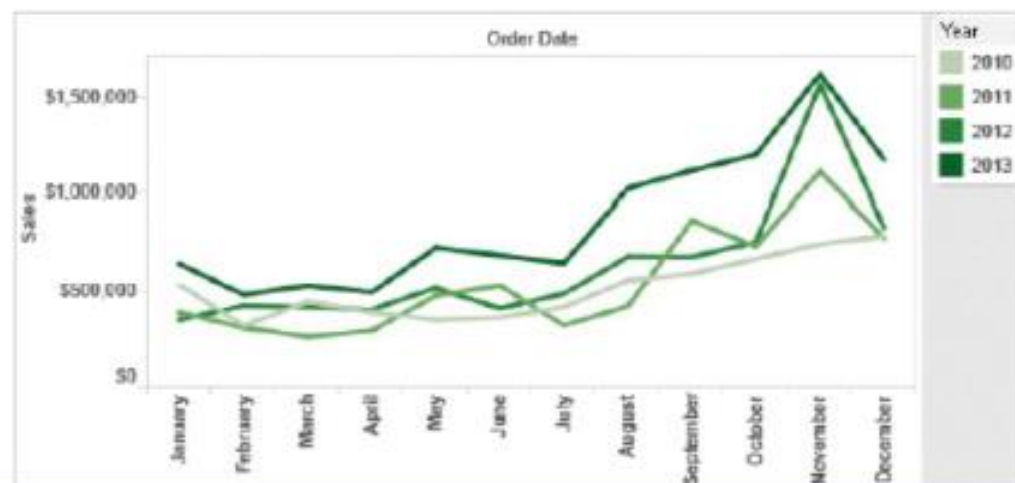
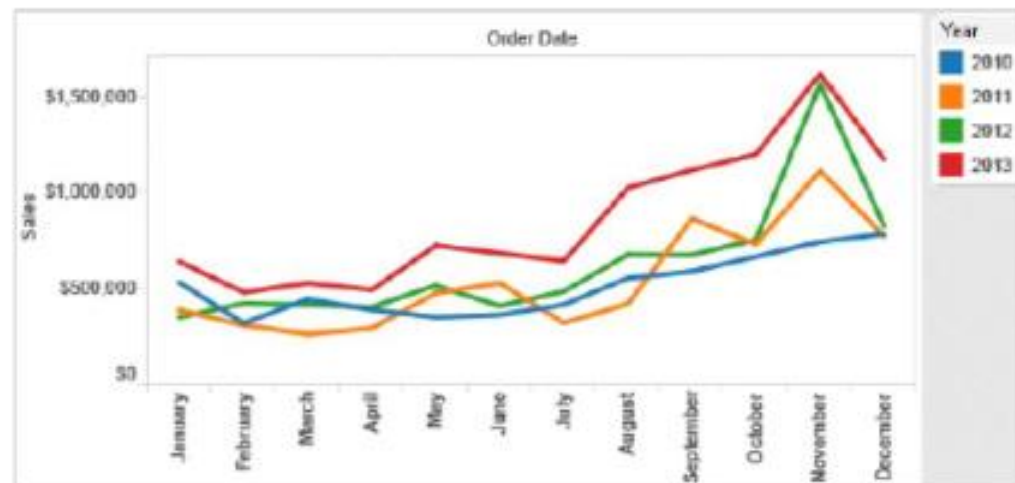


Task.

Which of the visualisations is easier to read?
Justify your answer.

5:00

5 Minute
Countdown Timer



Data Visualisation Packages.

Many companies offer downloadable tools that enable you to import data and to create reports and data visualisations from this data using a dedicated environment.

These software packages help users to manipulate their data at the click of a button, rather than forcing the user to write code.

Pros

Mostly no code needed

Simple user interface

Easier to learn

Cons

Limited functionality

Usually available at a cost

Less powerful

What's Out There?



alteryx

Tableau.



A public version that is free to run is available, this will enable you to practice with the software for free. Please do note that some of the features are limited to premium/paid for versions.

<https://public-pantheon.tableau.com/>

Dashboard Storytelling.

Data exploration of the Titanic dataset:

How many variables do we have in our dataset?

What type of variables the dataset has?

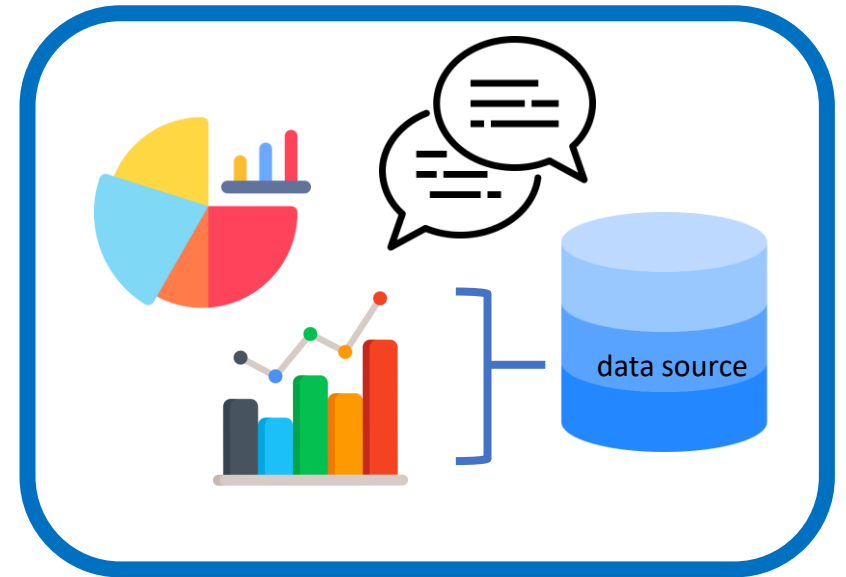
Questions to investigate:

What is the proportion of survivor by gender, and by age?

Which cabin class was the most affected?

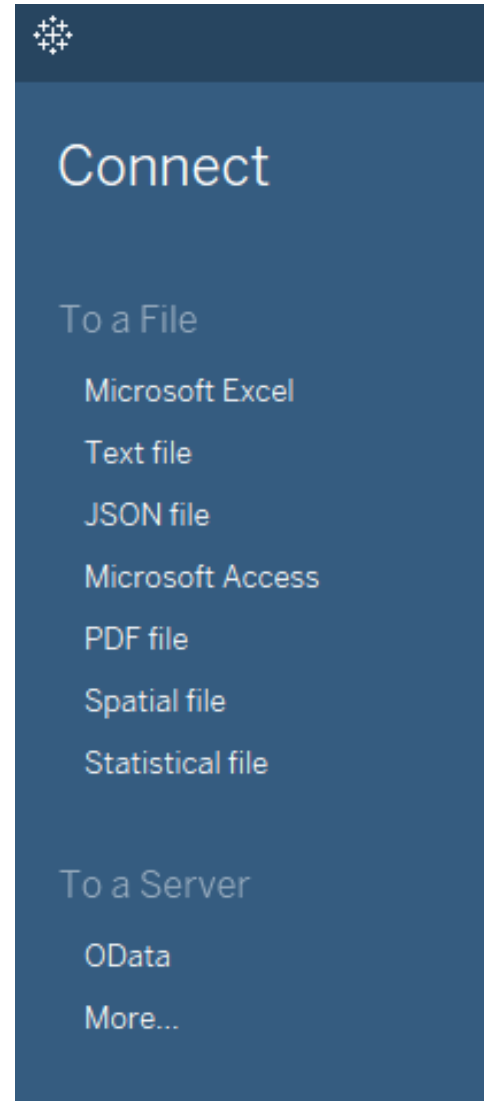
Make appropriate graphics, can add some narrative text

Combine data insights into a dashboard



Importing Data In Tableau.

- From the desktop public version of Tableau, you are able to import data from multiple locations.
- In this example we want to import our data from a CSV flat file.
- Select Text file and find the location of the saved CSV.



Data Exploration.

Data source
main page:

The screenshot shows the 'Data Source' main page for a connection named 'titanic train data set'. The interface includes a left sidebar with 'Connections' and 'Files' sections. The 'Connections' section lists 'titanic train data set' as a 'Text file'. The 'Files' section shows 'titanic train data set.csv' and a 'New Union' option. The main area displays a message 'Need more data? Drag tables here to relate them. [Learn more](#)'. Below this, a table summary shows '12 fields 891 rows'. A table preview is visible at the bottom, showing columns: Passenger Id, Survived, Pclass, Name, and Sex. The bottom navigation bar includes 'Data Source', 'Sheet 1', and 'Dashboard 1'.

Connections [Add](#)

- titanic train data set
Text file

Files [Add](#)

- ☐ Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.
- titanic train data set.csv
- New Union

titanic train data set

Connection: ☒ Live ☐ Extract Filters: 0 [Add](#)

titanic train data set.csv

Need more data?
Drag tables here to relate them. [Learn more](#)

titanic train data set.csv 12 fields 891 rows 100 rows [Settings](#)

| # | Passenger Id | Survived | Pclass | Name | Sex |
|---|------------------------------------|----------|--------|--------|-----|
| 1 | Braund, Mr. Owen Harris | 0 | 3 | male | |
| 2 | Cumings, Mrs. John Bradley (...) | 1 | 1 | female | |
| 3 | Heikkinen, Miss. Laina | 1 | 3 | female | |
| 4 | Futrelle, Mrs. Jacques Heath (...) | 1 | 1 | female | |
| 5 | Allen, Mr. William Henry | 0 | 3 | male | |
| 6 | Moran, Mr. James | 0 | 3 | male | |
| 7 | McCarthy, Mr. Timothy J | 0 | 1 | male | |
| 8 | Palsson, Master. Gosta Leonard | 0 | 3 | male | |

Fields

| Type | Field Name | Physical Table | Remote... |
|------|--------------|----------------------------|------------|
| # | Passenger Id | titanic train data set.csv | Passeng... |
| # | Survived | titanic train data set.csv | Survived |
| # | Pclass | titanic train data set.csv | Pclass |
| Abc | Name | titanic train data set.csv | Name |
| Abc | Sex | titanic train data set.csv | Sex |

Data Source Sheet 1 Dashboard 1

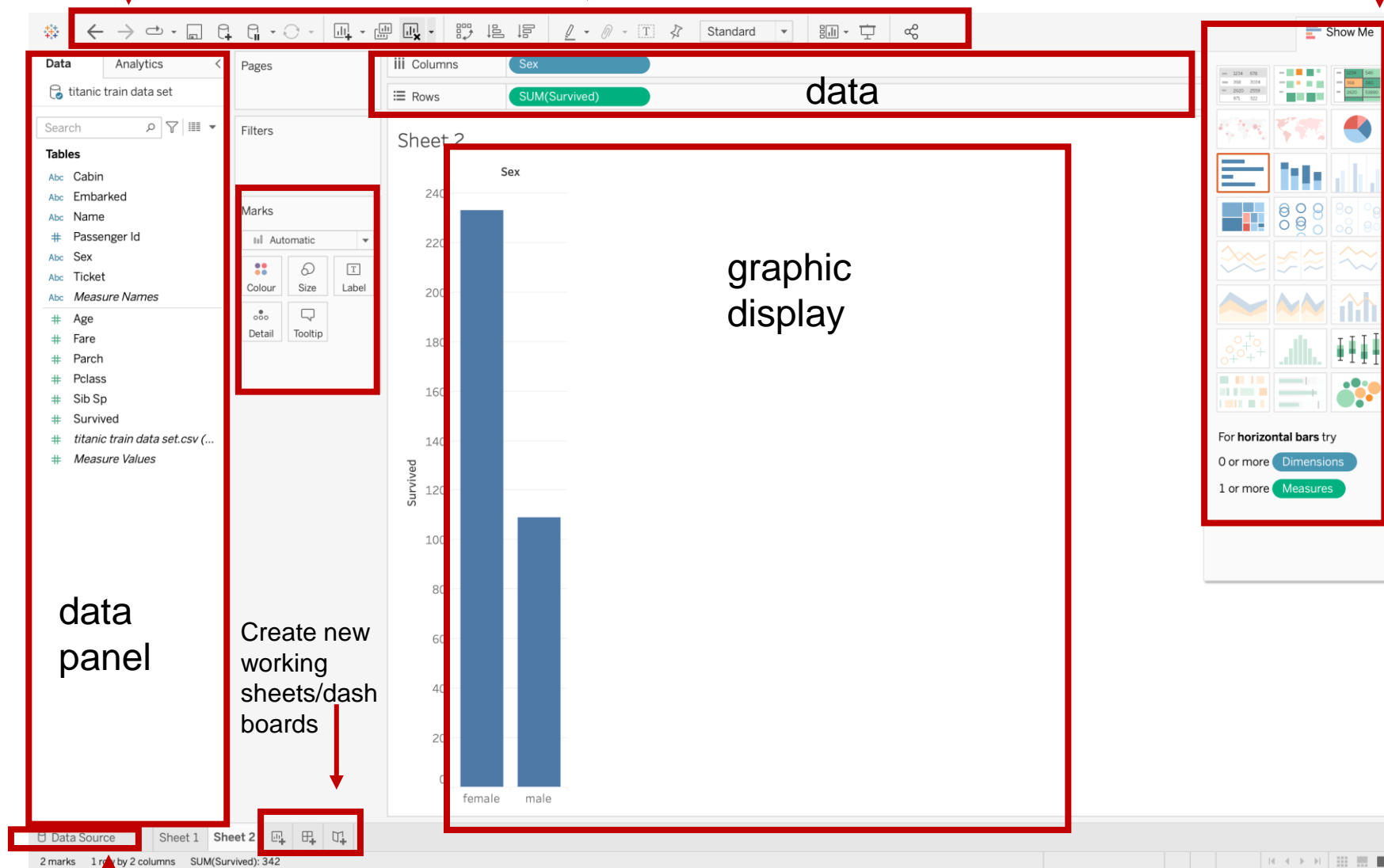
Explore
the type
of data

Data Exploration.

undo and redo

data sorting

graphic type



Go back to the initial data source

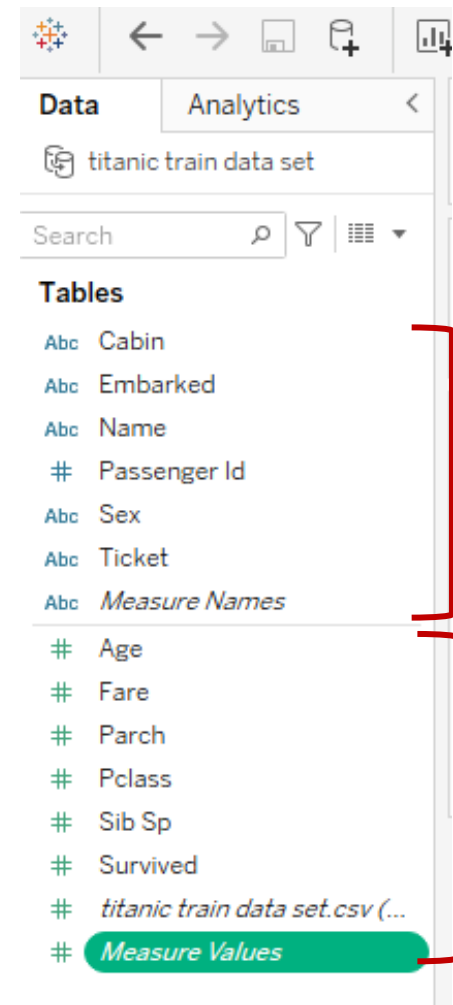
Data Exploration.

When data is imported, Tableau will automatically decide if the variables are dimension or measures.

Dimensions = categories e.g. race, sex, and educational level.

Measures = data type expressed in numbers e.g. Age, height, and length.

A variable can be converted into a dimension or measure if necessary.



The screenshot shows the Tableau 'Data' pane for the 'titanic train data set'. The variables are listed in two groups:

- Dimensions:** Cabin, Embarked, Name, Passenger Id, Sex, Ticket, Measure Names.
- Measures:** Age, Fare, Parch, Pclass, Sib Sp, Survived, titanic train data set.csv (...), Measure Values.

Red brackets on the right side of the list group the variables into 'dimensions' and 'measures'.

Which Charts Are The Best To Plot My Data?

Some examples:

Bar charts and pie charts are best to plot categorical variables (race, sex, and educational level)

Data relationship/correlation can be visualised with a scatter plot or a heatmap

Line charts are best for plotting time series data that contain date time data over a period of time

Some resources: <https://medium.com/@wenjunwu/60-charts-to-visualize-your-data-51344d7178d1>
<https://www.kaggle.com/getting-started/160583>



Data Analysis.

Let's make our first graph to feed our dashboard answering the following question:

Q1: What is the proportion of Titanic survivor by gender?

To investigate whether men or women were the most affected by the sink of the Titanic, we can gather from our dataset the variable "Sex" which gives the two values female and male associated to each passenger ID.

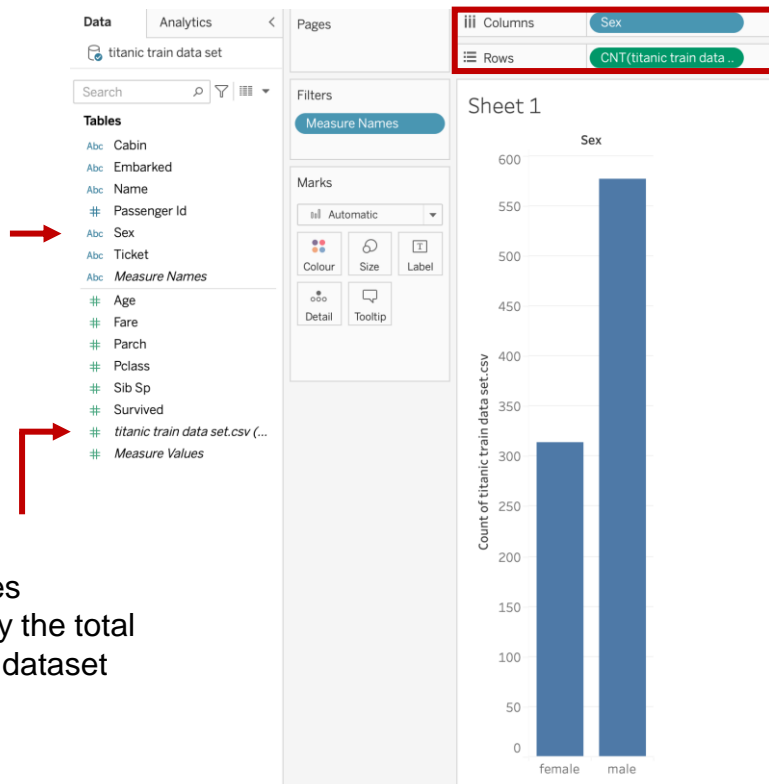
Additionally, Tableau has an automatic feature which gives the total number of records from a dataset; Using this feature we will be able to sort the total number of men and women in the dataset and determine how many have survived using the variable "Survived".

Data Analysis.

We will use a grouped bar chart that seems to be a good choice for comparing categorical variables

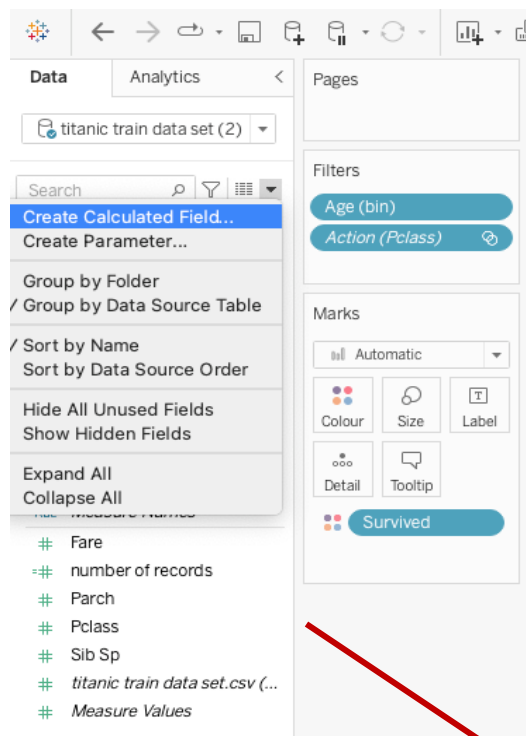
Drag the variable **Sex** into the Columns section (blue), and drag the **total count of the dataset** into the Rows section (green)

Right click on the **total count of the dataset (CNT)** to calculate the proportion of male and female in %



Data Analysis.

Create a calculation:



You can also create a calculated field which will automatically calculate the SUM of Records by right clicking in the space under the measure names > create calculated field.

It is possible to drag the variables name in the calculation box directly

Number of records

Change the name of the calculated field.
In the text box right the number 1 > click apply > click ok.

The calculation is valid.

Apply

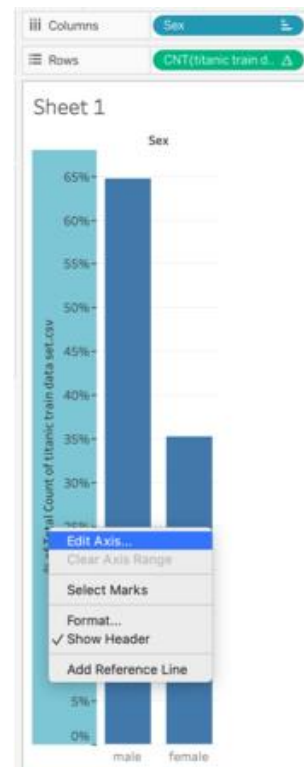
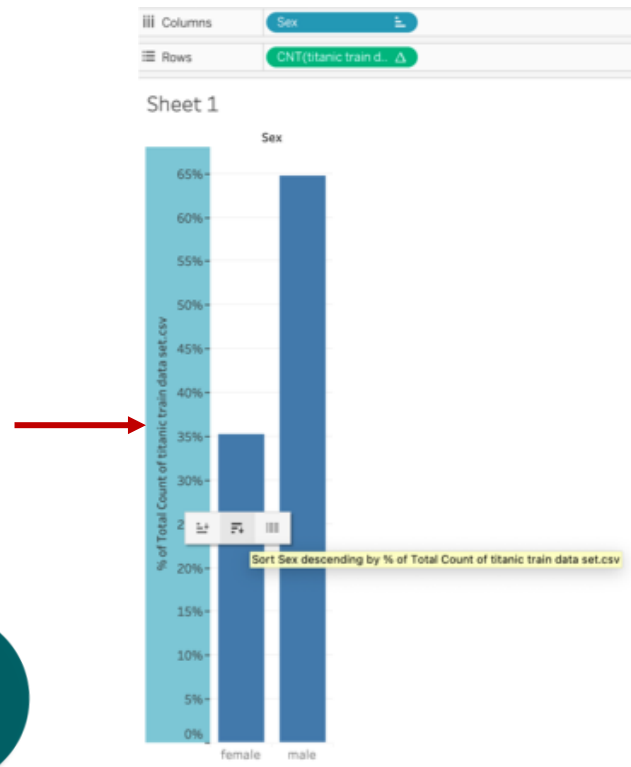
OK

Data Analysis.

Click on the Y axis to sort
your data in
ascending/descending order

To rename the Y axis right
click on the axis label and
select **Edit Axis**

pop up
window



Sex

CNT(titanic train d. ...)

Edit Axis [% of Total Count of titanic train data set.csv]

General

Range

☒ Automatic ☐ Uniform axis range for all rows or columns ☐ Independent axis ranges for each row or column ☐ Fixed

Automatic Automatic

0 0.079966330

Scale

☐ Reversed ☐ Logarithmic

☒ Positive ☐ Symmetric

Axis Titles

Title

Proportion of male and female in the titanic (%)

Subtitle

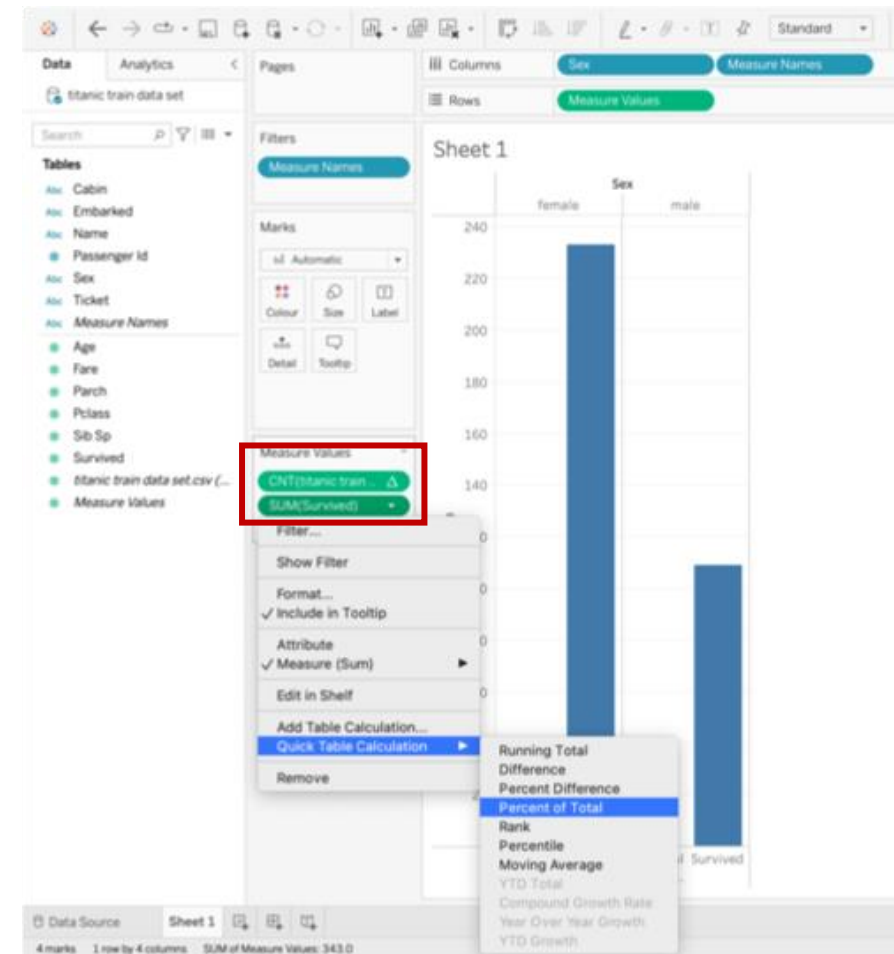
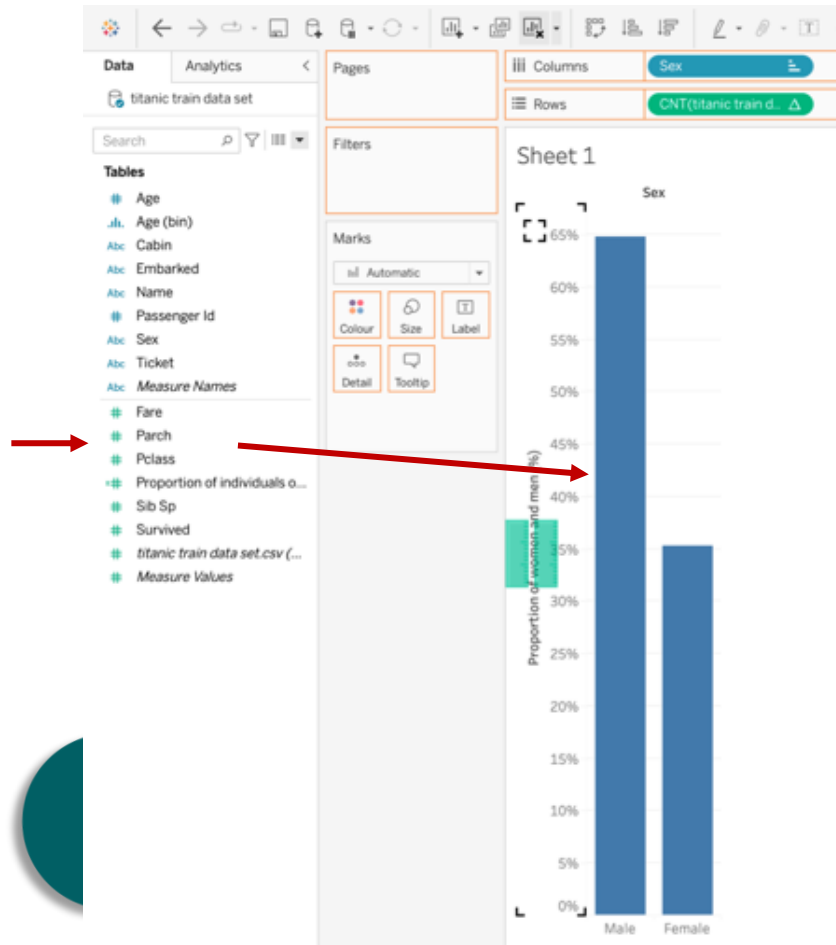
Automatic

Reset

Data Analysis.

Drag and drop the **Survived** variable in the Y axis of the graph

This will create a new value in the Measure Values section. Transform this value in % by right clicking on **SUM (Survived)**

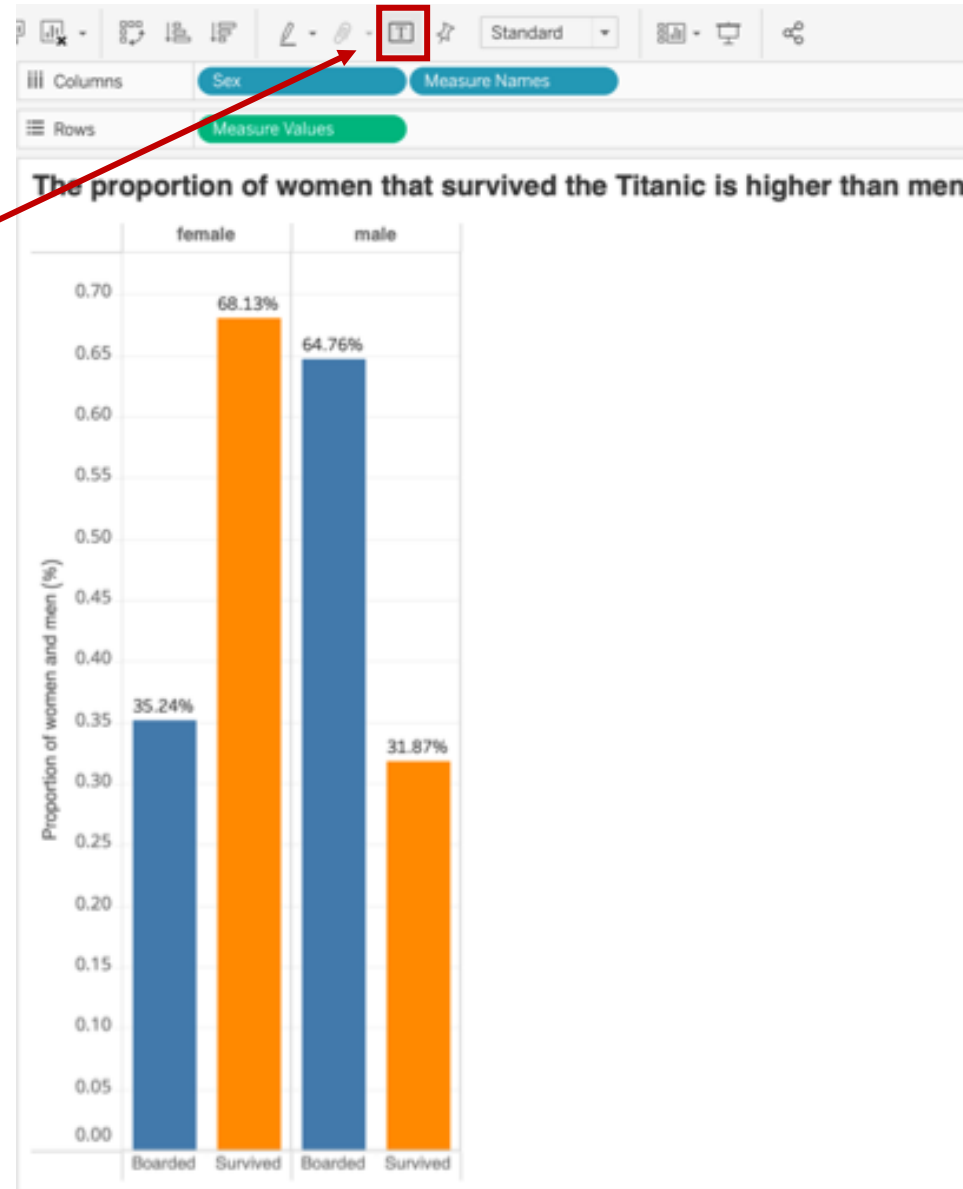


Data Analysis.

You should get the following graph.

Add a title by double clicking on top of the graph in the sheet section.

Display % on the top of each bar

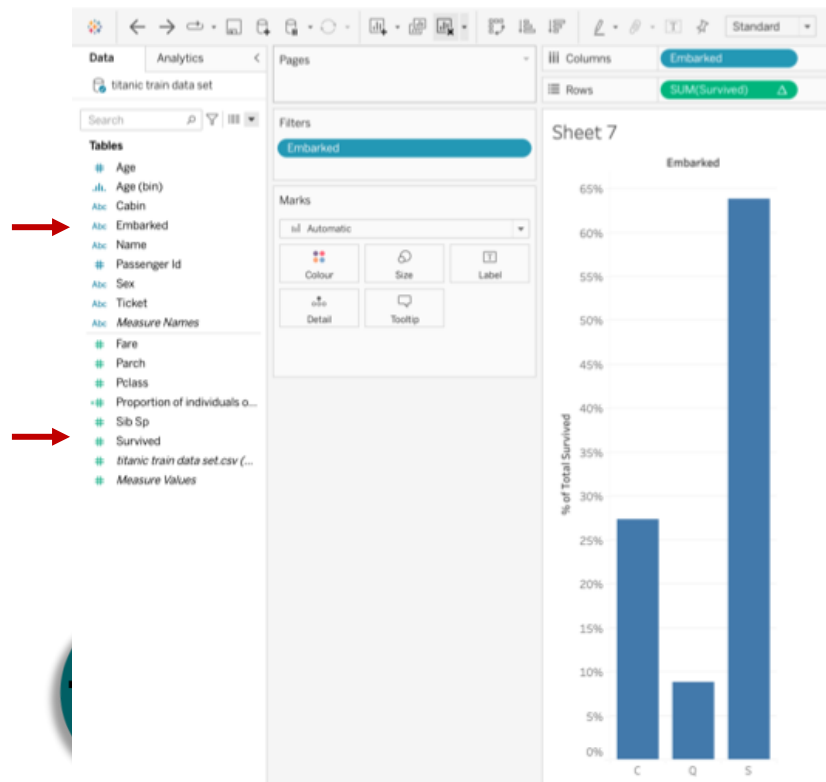


Data Analysis.

Q2: Which cabin class has the most survivor? =>

Create a new sheet

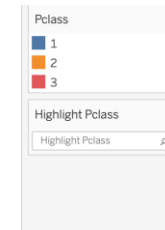
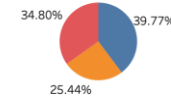
Select the **Embarked** and **Survived** variables



change chart type to
pie chart

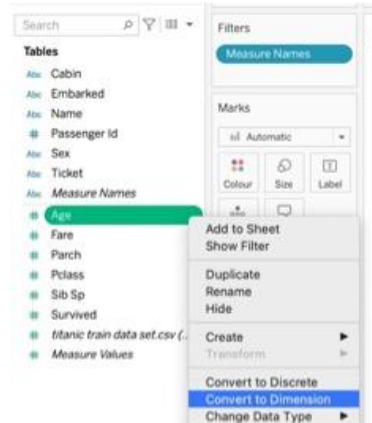


Passengers from the 1st class were less affected during the sink of the Titanic



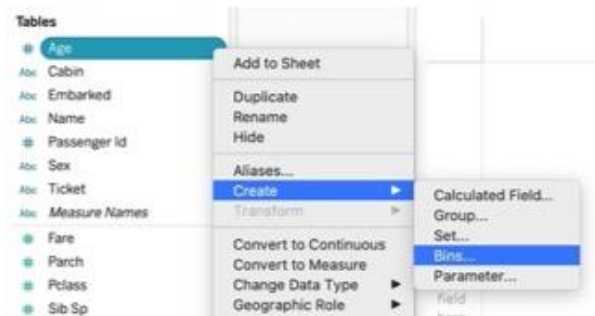
Data Analysis.

Q3: What is the proportion of survivors by age category?



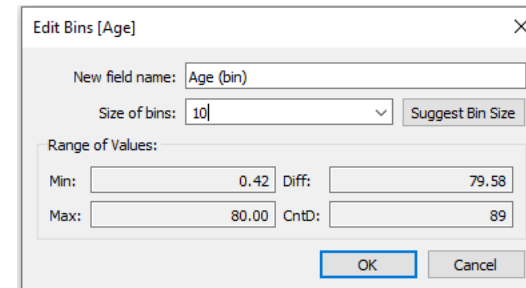
Converting a measure into a dimension:

Right click on the variable name >
Convert to Dimension



Divide the age variable into bins:

Change the name and the size of the bin if necessary (choose bin of 10)



Tables

- .ili. Age (bin)
- Abc Cabin
- Abc Embarked
- Abc Name
- # Passenger Id
- Abc Sex
- Abc Ticket

It has now been converted

Null Values.

Ctrl and left click to drag and drop the variable name into the filter box.

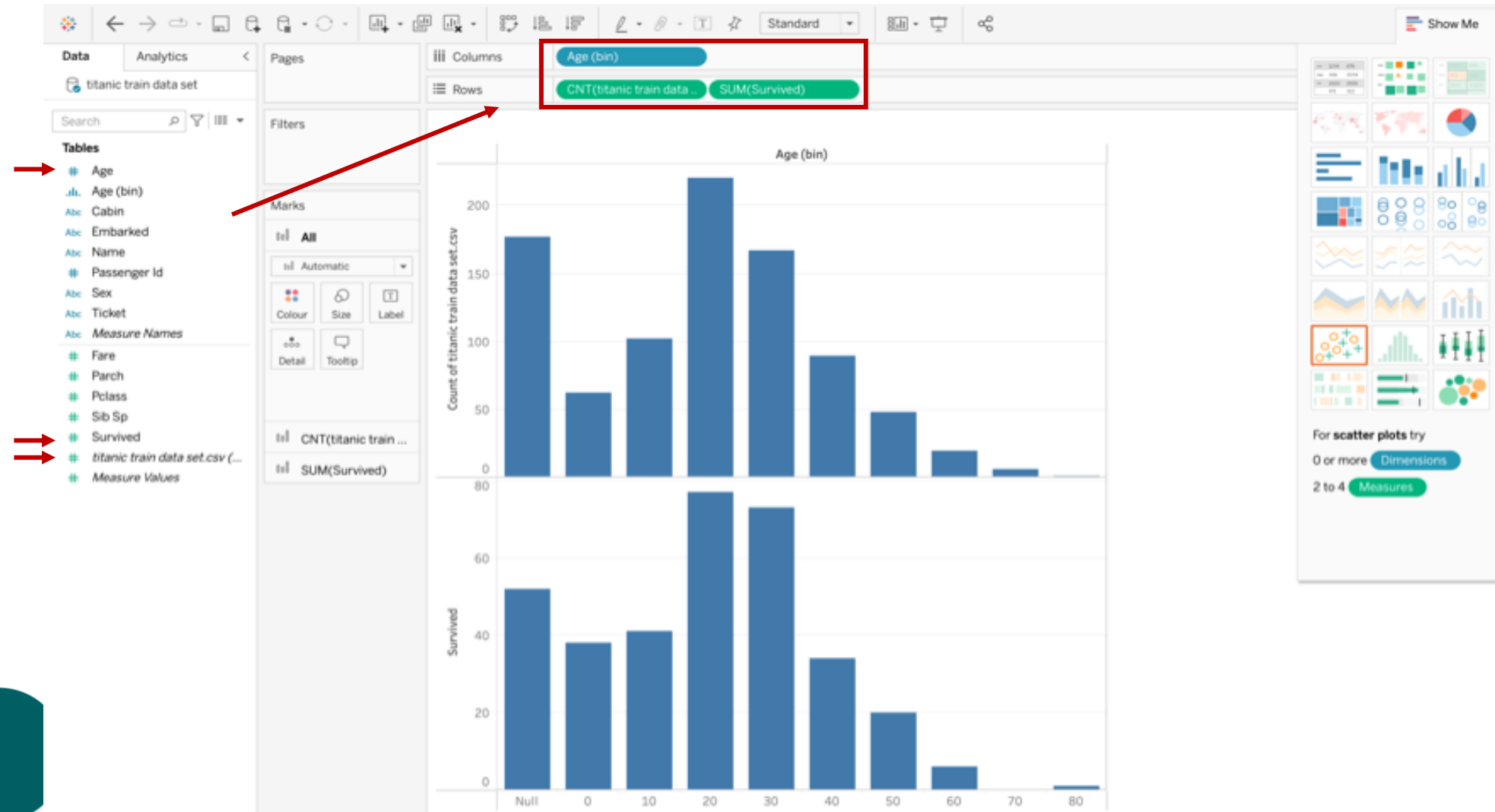
From the list of variable values make sure the Null box is NOT selected.

This will ensure Null values from this bin will not be included.

The screenshot shows the Tableau interface with a bar chart titled 'Sheet 1'. The Y-axis is labeled 'Age' and ranges from 0K to 5K. The X-axis is labeled 'Age (bin)'. A red arrow points from the 'Age (bin)' variable in the Columns shelf to the 'Filter [Age (bin)]' dialog box. The dialog box has three tabs: 'General', 'Condition', and 'Top'. The 'General' tab is selected. It shows a list of values for 'Age (bin)' with checkboxes. The 'Null' checkbox is unchecked, while all other values (0, 10, 20, 30, 40, 50, 60, 70, 80) are checked. The 'Summary' section shows 'Field: [Age (bin)]', 'Selection: Selected 9 of 10 values', 'Wildcard: All', 'Condition: None', and 'Limit: None'. The 'All' and 'None' buttons are visible, along with an 'Exclude' checkbox. The 'OK', 'Cancel', 'Apply', and 'Reset' buttons are at the bottom.

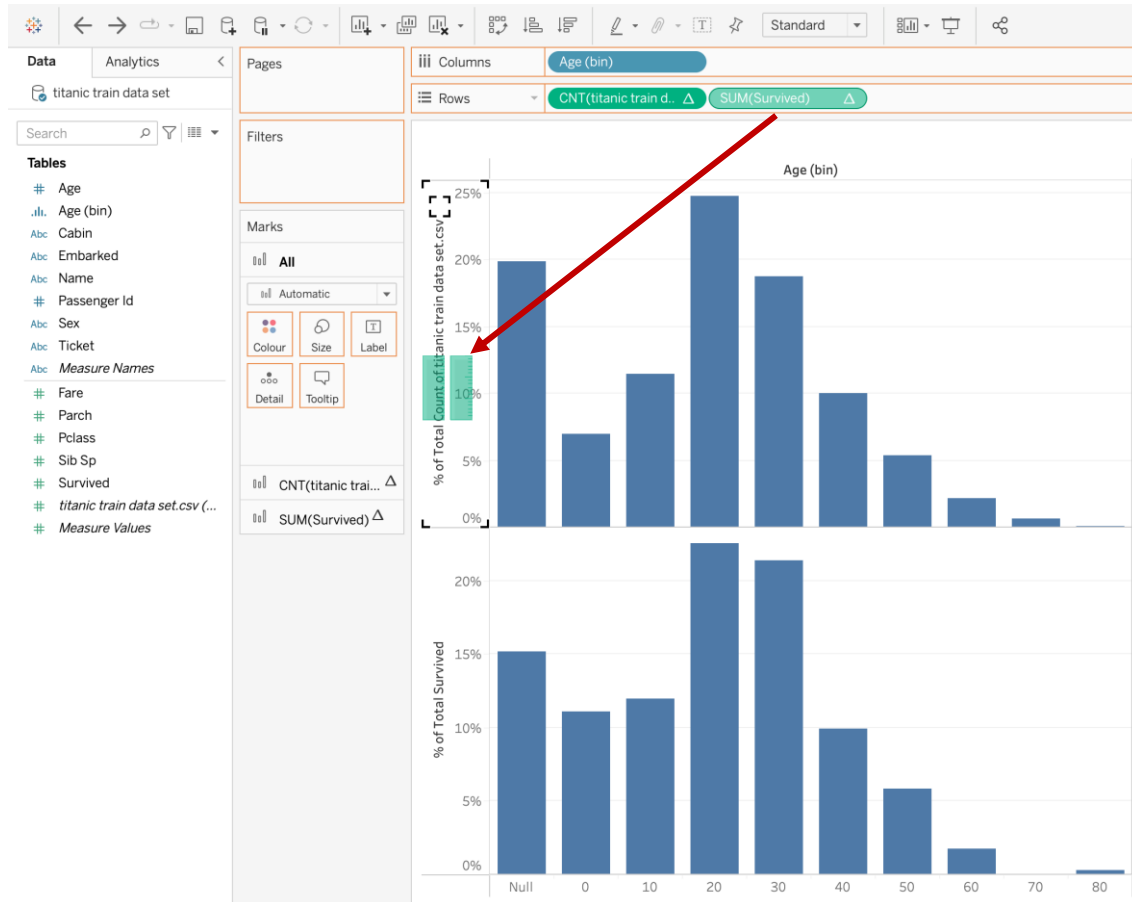
Data Analysis.

Drag and drop corresponding variables, and convert them in percentage

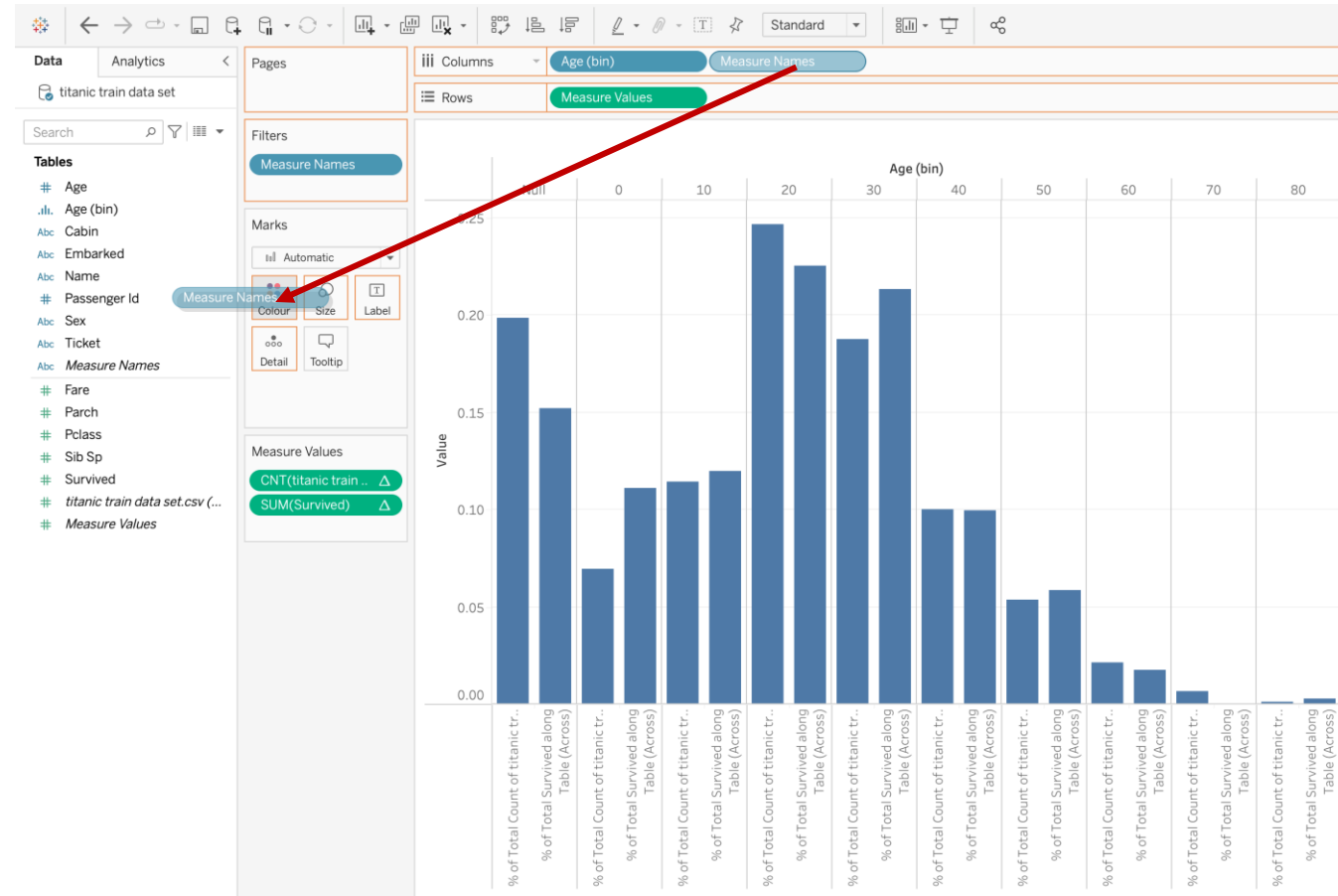


Data Analysis.

drag and drop SUM(survived) into the Y axis



drag and drop Measure Names into colours



Data Analysis.

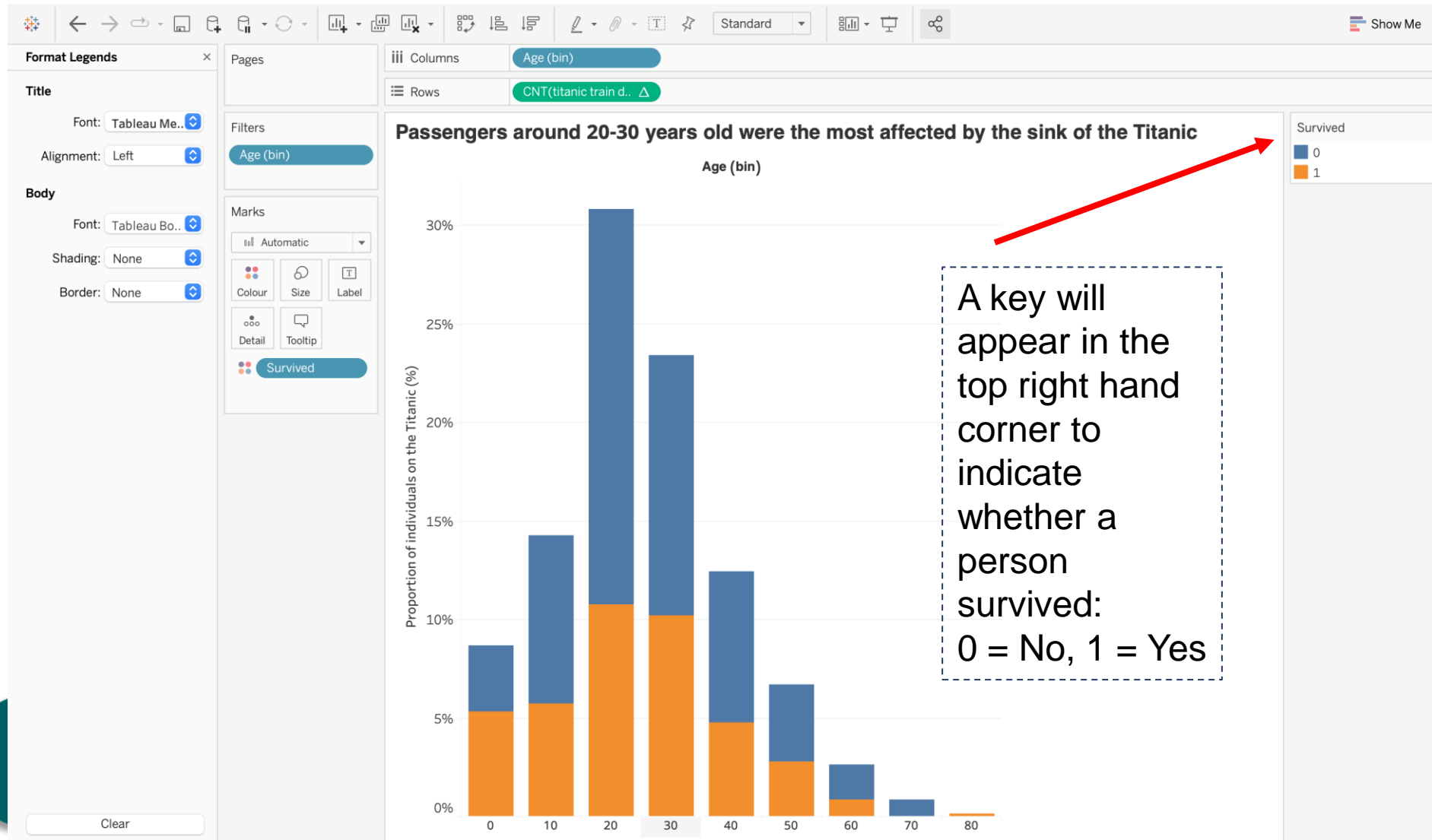
Convert the Survived variable to Dimension

The screenshot shows the Tableau 'Data' pane on the left. Under the 'Tables' section, the 'Survived' variable is highlighted with a red arrow. A context menu is open over 'Survived', and the option 'Convert to Dimension' is selected and highlighted in blue. Other options in the menu include 'Add to Sheet', 'Show Filter', 'Duplicate', 'Rename', 'Hide', 'Create', 'Transform', 'Convert to Discrete', 'Change Data Type', 'Geographic Role', 'Default Properties', 'Group by', and 'Folders'.

Drag Survived into the colour section

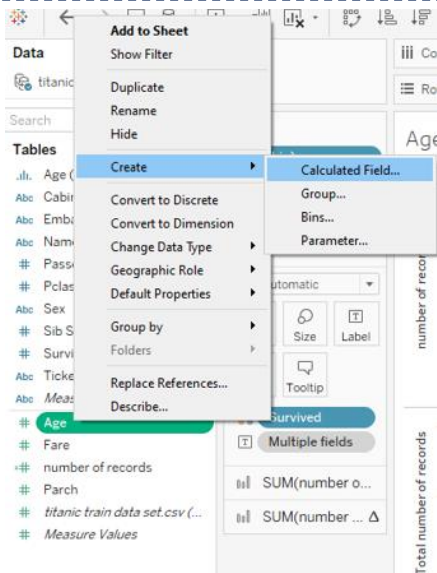
The screenshot shows the Tableau 'Data' pane on the left. The 'Survived' variable has been dragged from the 'Tables' list to the 'Colour' property of the 'Marks' card. A red arrow points from the 'Survived' variable in the 'Tables' list to the 'Colour' property in the 'Marks' card. The 'Measure Values' section now shows two measures: 'CNT(titanic train dat..)' and 'SUM(Survived)'.

Data Analysis.

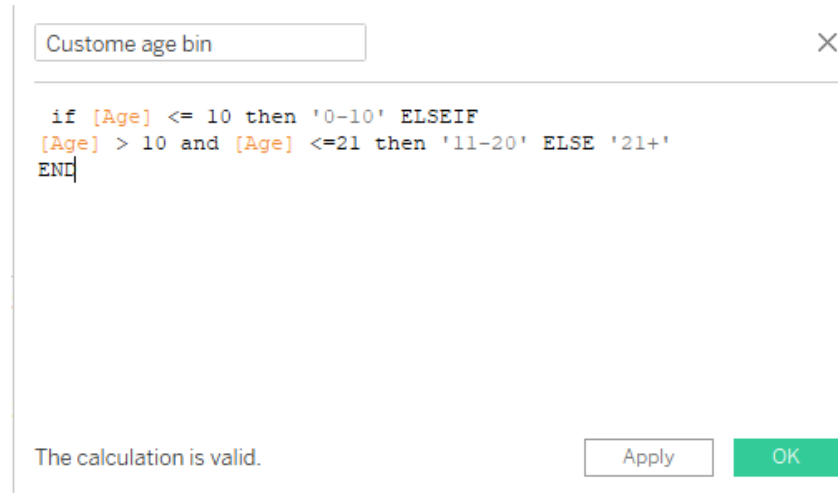


Custom Bins.

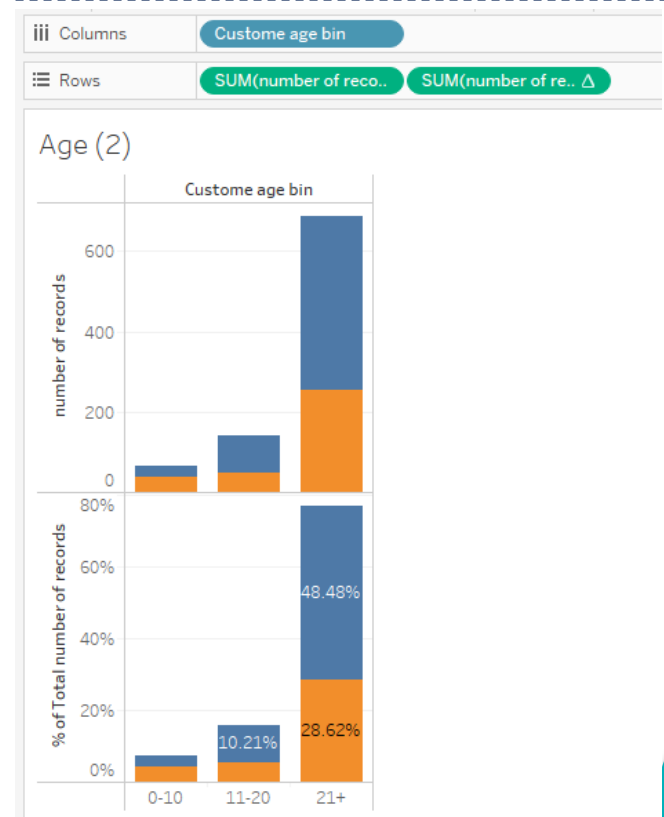
Right click on bin name >
create > calculated field



Write an IF statement to
set the parameters

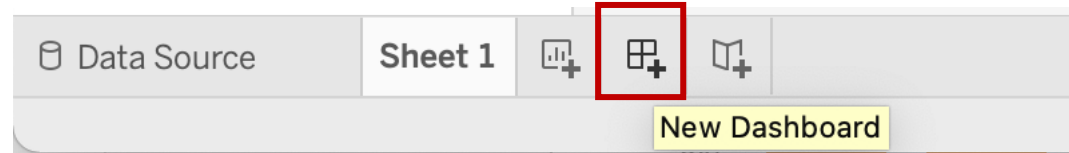


Drag and drop the new
customised bin into the column
section

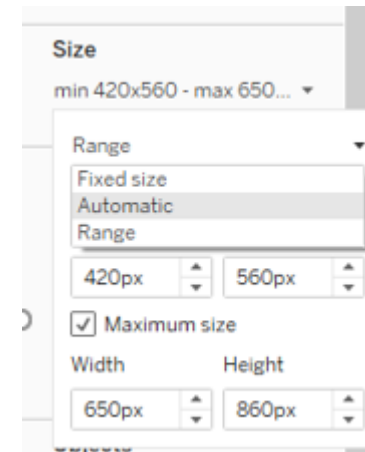


Creating a Dashboard.

Select New Dashboard tab at the bottom of the page

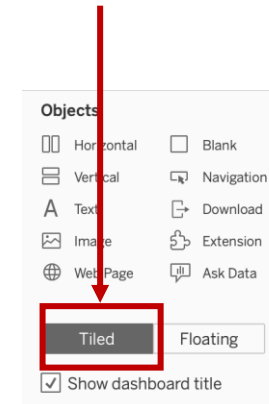


Change size to automatic



Use the Tiled method to dispose your sheets into the dashboard initially

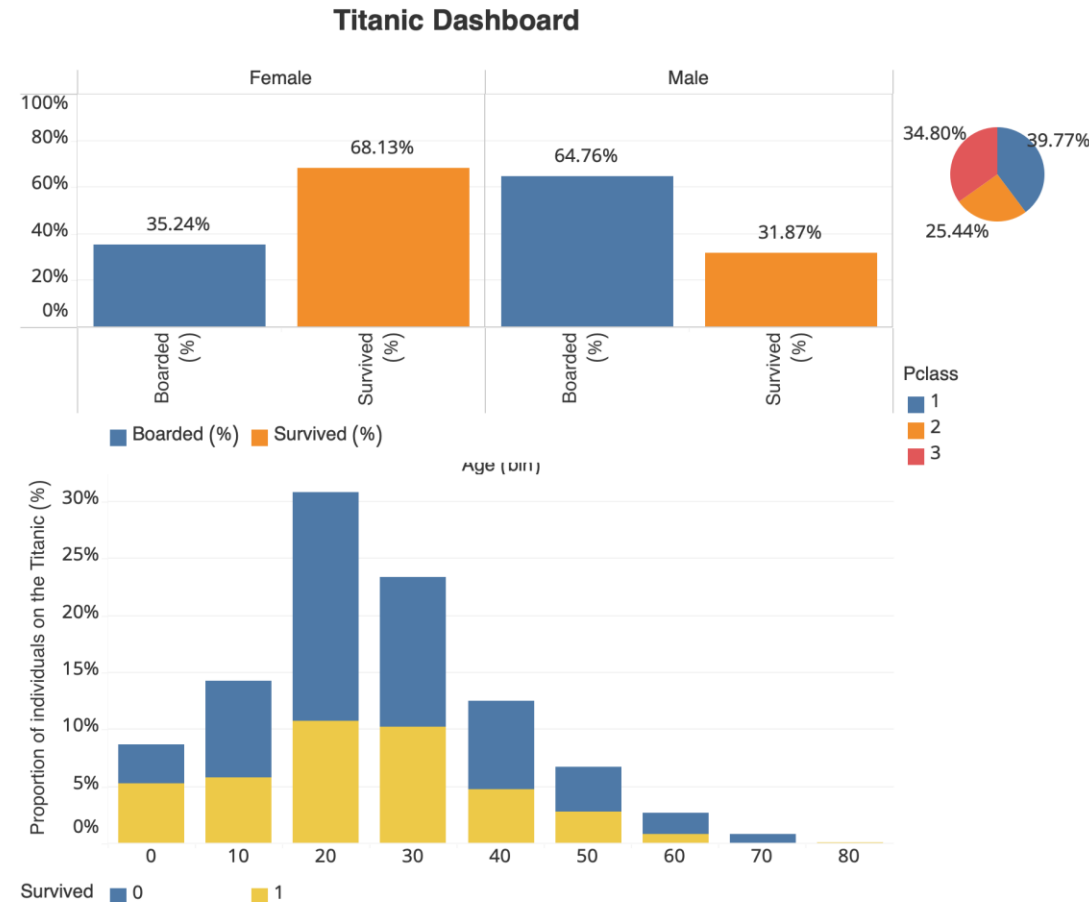
Double click on the sheets you want to view on your dashboard



Dashboard.

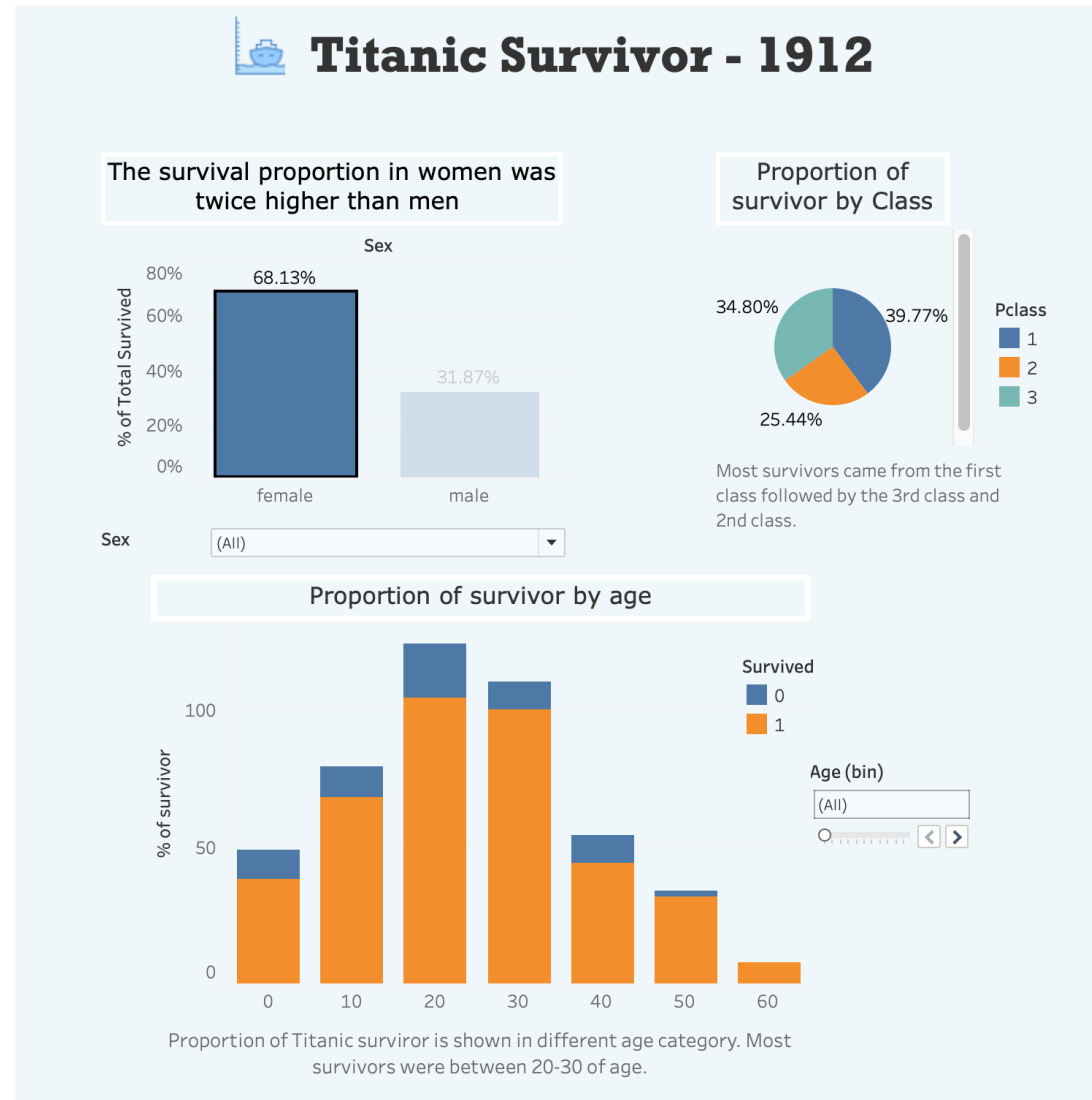
Here you can change the size of each individual graphs by dragging the sheets into the dashboard. You should have 3 sheets with graphs to make your dashboard.

On top of tableau, click `format>Workbook:` you can change the font and size of the all dashboard



Keep your fonts at minimum for consistency: max 2 fonts (one for the title and one for the rest)

Dashboard.



Additional Resources.

Choosing the right colours:

<https://www.storytellingwithdata.com/blog/2020/5/6/picking-the-right-colors>

Colour theory:

https://tools.picsart.com/color/wheel/?utm_source=google&utm_medium=ppc&utm_campaign=15919706814&adgroupid=131972401123&utm_term=a%20color%20wheel&gclid=Cj0KCQjwzLCVBhD3ARIsAPKYTcTpGhNTowoMhXFPup2T1srfsyn0XqUTQyFwohaKK8gETjflGoY5EqAaAvEyEALw_wcB

Choosing the right fonts:

<https://medium.com/nightingale/choosing-a-font-for-your-data-visualization-2ed37afea637>

Tableau Visual best practices:

https://help.tableau.com/current/pro/desktop/en-us/visual_best_practices.htm





Plenary.

1. What makes a good data visualisation?
2. How do you use colour in your visualisations?
3. What is a scatter plot? What types of data work best in scatter plots?