Universidad de La Laguna

Data Visualization with Chart.js

Sara Revilla Báez
Francisco Javier Arocas Herrera

Table of contents

1.

Data Visualization

An introduction

What is Data Visualization?

Data visualization is the graphic representation of data.

The importance of Data Visualization

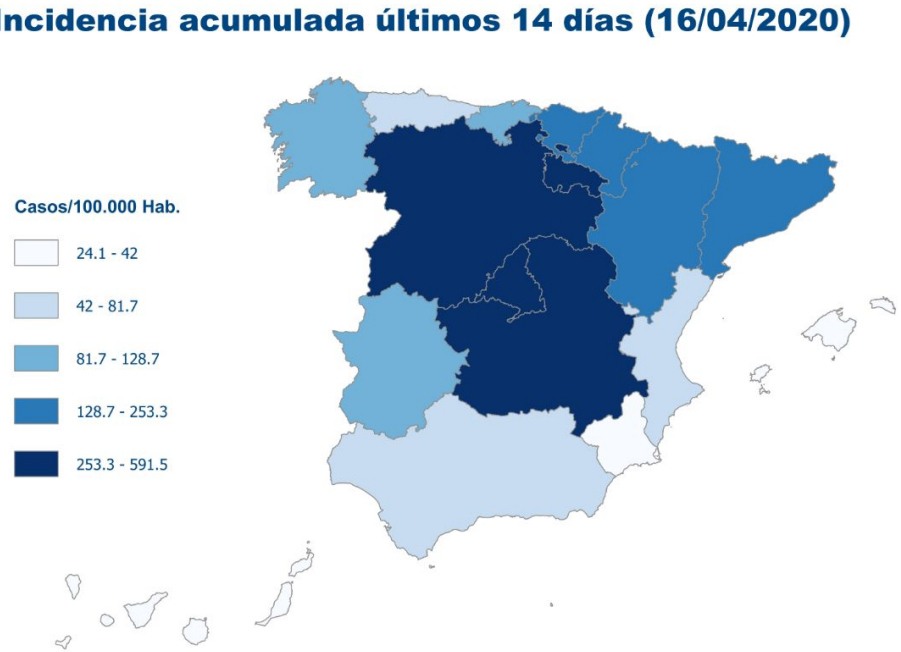
People are not equipped to understand large amounts of words and numbers.

Casos por CCAA

CCAA	Total	Ult.24h	Inc.14d
Andalucía	11053	246	43.72
Aragón	4664	98	134.54
Principado de Asturias	2230	60	77.92
Islas Baleares	1668	31	35.76
Canarias	2009	21	24.10
Cantabria	1884	39	96.89
Castilla y León	14903	523	292.89
Castilla La Mancha	15997	846	367.66
Cataluña	38316	962	193.56
Galicia	8013	140	103.50
C. Valenciana	9869	254	64.85
Extremadura	3019	138	105.46
Comunidad de Madrid	51993	1299	267.21
Región de Murcia	1625	27	32.13
Comunidad Foral de Navarra	4433	85	244.11
País Vasco	12089	299	193.04
La Rioja	4098	182	591.54
Ceuta	101	1	46.00
Melilla	104	1	39.31

Source: covid19.isciii.es (17/04/2020)

The importance of Data Visualization



Source: covid19.isciii.es (17/04/2020)

A picture is worth

*a **thousand** words.*

The importance of Data Visualization

Visualization makes it easier to analyze complex data and extract information from it.

- To choose the right visualization we must understand the type of data we're working with.

Types of Data

Qualitative (Categorical)

Quantitative (Numerical)

- Continuous
- Discrete

Temporal

Spatial

Example: COVID-19 data

- Number of confirmed cases
- Country
  - Region
- Date

Example: COVID-19 data

- Number of confirmed cases → Cuantitative (Discrete)
- Country
  - Region
- Date

Example: COVID-19 data

- Number of confirmed cases → Cuantitative (Discrete)
- Country → Spatial
  - Region
- Date

Example: COVID-19 data

- Number of confirmed cases → Cuantitative (Discrete)
- Country → Spatial
  - Region
- Date → Temporal

What makes a good visualization?

A graph should:

- Be self-explanatory
- Show the data without distorting it
- Induce the viewer to think

What makes a good visualization?

A graph should:

- Be self-explanatory

  → Use a legend, labels for the axes and titles
- Show the data without distorting it

- Induce the viewer to think

What makes a good visualization?

A graph should:

- Be self-explanatory

  → Use a legend, labels for the axes and titles
- Show the data without distorting it

  → Adjust the scales, use color, shape and size
- Induce the viewer to think

Keep it simple.

2.

Chart.js

Installation and use

What is Chart.js?

Open source JavaScript library, simple and flexible graphics for designers and developers.

- Widely used
- Interactive graphs
- Responsive
- 8 Chart types

Getting started

Using NPM:

1. Installation:

```
npm install chart.js
--save
```

2. Use:

```
<script
src="YOUR_PATH/chart.j
s"></script>
```

Using a CDN:

```
<script
src="https://cdn.jsdel
ivr.net/npm/chart.js@2
.8.0"></script>
```

Drawing a chart

1. Create a Canvas, where the chart will be drawn
2. Specify the type of chart to draw
3. Supply the parameters:
    a. Data
    b. Labels
    c. Options

Drawing a chart

```
let ctx =
document.getElementById('myChart').getContext('2d');
let chart = new Chart(ctx, {
    type: 'line',     // The type of chart
    data: {           // The data
    ...
    },
    options: {}       // Configuration options
});
```

Data

```
data: {
    labels: ['First', 'Second', ...],
    datasets: [
        {
            label: 'First Dataset',
            data: [...],
            property: value
        },
        { label: 'Second Dataset', data: [...] }
    ]
}
```

Options

- Title
- Legend
- Scales

3.
Plots

Types, uses, tips and examples

Plots

Which visualization is best for showing certain information from a given dataset?

- Comparison plots
- Relation plots
- Composition plots
- Distribution plots
- Geoplots

Comparison Plots

Used for comparing values over time or between different categories or variables.

1. Line Chart
2. Bar Chart
3. Radar Chart

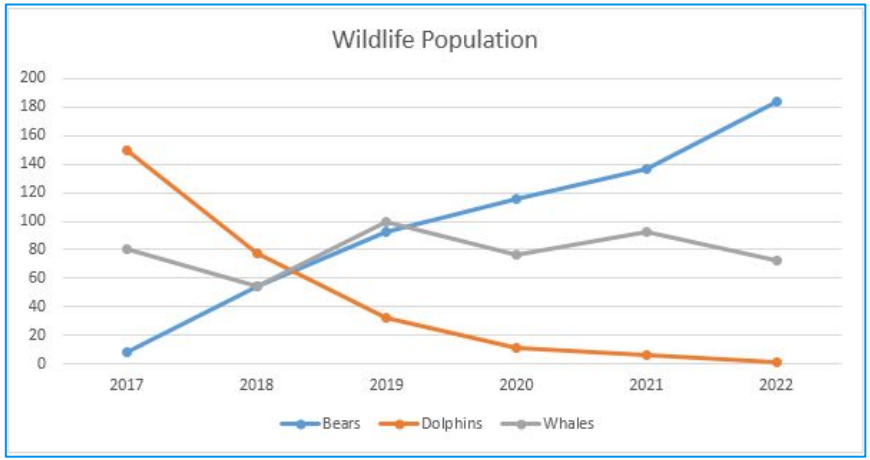
Line Chart

BEST FOR: Visualizing one or multiple variables over time (10+ days)

TIPS:

- Avoid too many lines per chart
- Adjust the scale adequately
- Use labels for multiple variables

Wildlife Population between 2017 and 2020



Bar Chart

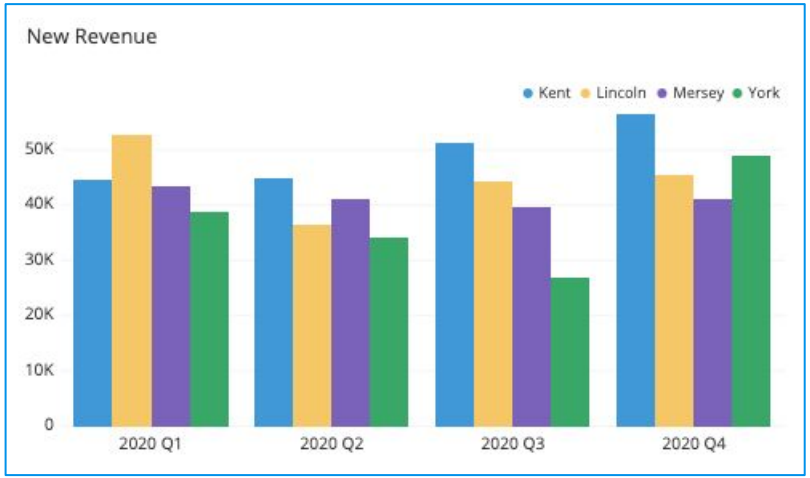
BEST FOR: Comparing categories or a single variable over a short period of time. They can be vertical or horizontal.

TIPS:

- The axis for the numerical variable should start at zero
- Use labels

DON'T CONFUSE IT WITH A HISTOGRAM

Shop Benefits in 2020 (divided into quarters)



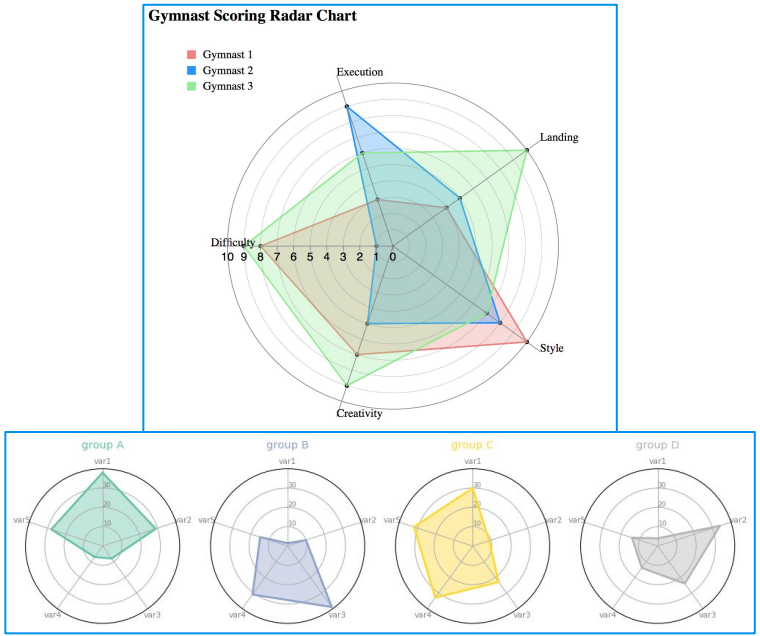
Radar Chart

BEST FOR: Multiple quantitative variables. Each variable is plotted on its own axis.

TIPS:

- Display 10 variables or less to make it easier to read
- Use **faceting** for different groups/each variable

3 Gymnast Scoring Values



Relation Plots

Used for detecting relationships between variables.

1. Scatter Plot
2. Bubble Plot
3. Correlogram
4. Heatmap

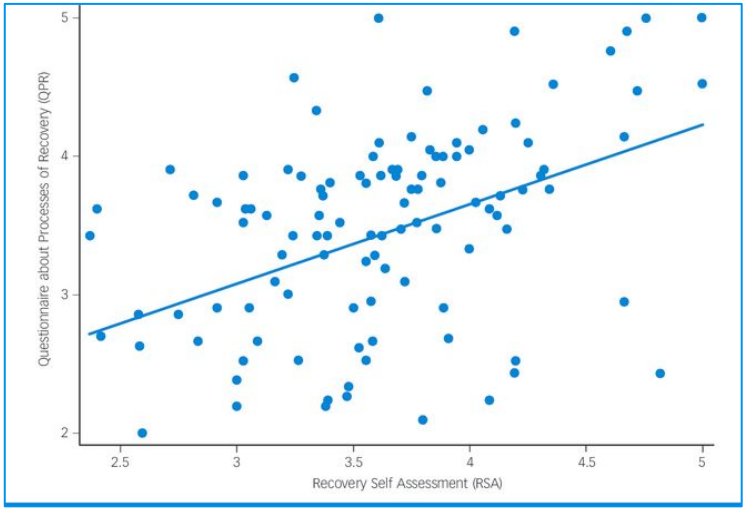
Scatter Plot

BEST FOR: Detecting a correlation between two variables

TIPS:

- Both axes should start from zero
- Use color to represent multiple groups or categories
- Marginal histograms can be added

Relationship between (RSA) and (QPR)



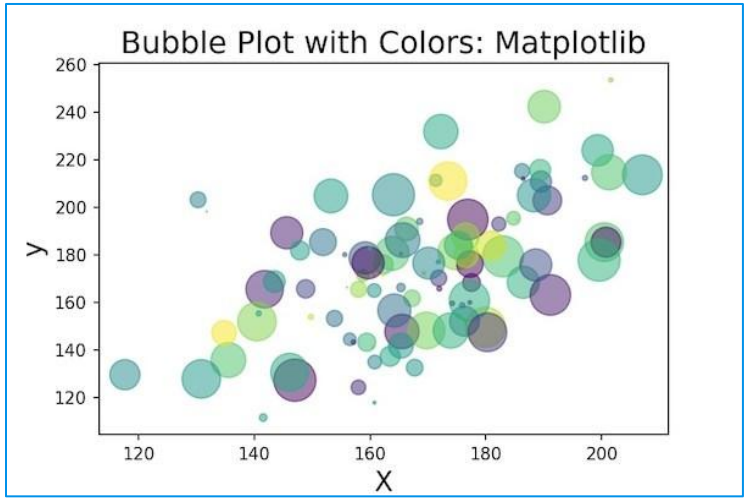
Bubble Plot

BEST FOR: Detecting correlations between **THREE** variables. It's a scatter plot with dots of different sizes.

TIPS:

- Both axes should start from zero
- Use color to represent multiple groups or categories
- Don't use it for large amounts of data, it makes the reading difficult

Bubble Plot generated using MatPlotlib (Python)



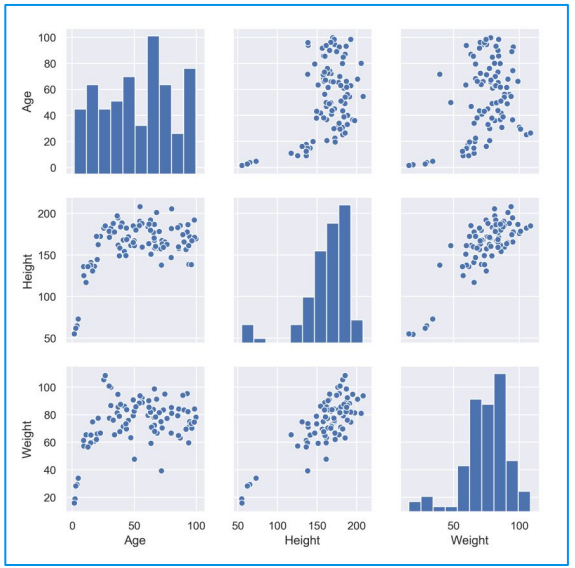
Correlogram

BEST FOR: Detecting the correlation between multiple variable pairs. It's a combination of scatter plots and histograms.

TIPS:

- Both axes should start from zero
- Use color to represent multiple groups or categories

Relationships between age, height and weight



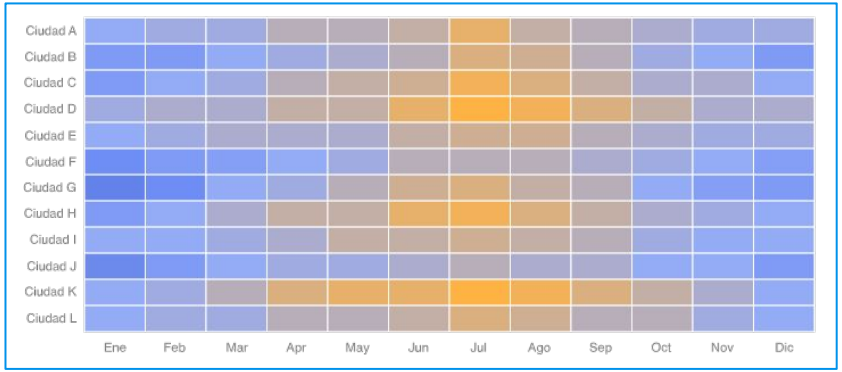
HeatMap

BEST FOR: Finding patterns in multivariate data

TIPS:

- Use contrasting colors
- Add a legend that shows the value each color corresponds to

Monthly temperature in cities



Composition Plots

Used to illustrate proportions or subcategories.

1. Pie Chart (Donut Chart)
2. Stacked Bar Chart
3. Stacked Area Chart

Pie Chart (Donut Chart)

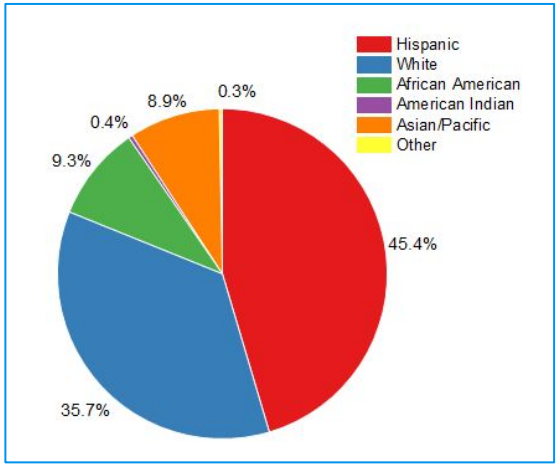
BEST FOR: Comparing items that are part of a whole or proportions

TIPS:

- Consider using a bar chart
- Arrange the slices according to their size
- Use contrasting colors

VARIANT: Donut Chart

Percentage of population



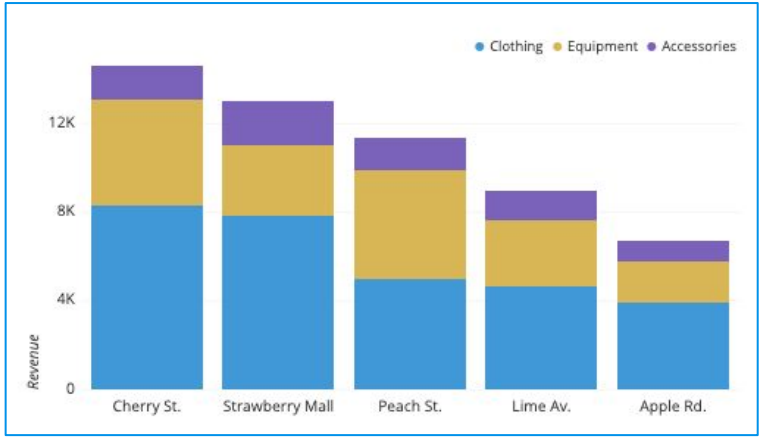
Stacked Bar Chart

BEST FOR: Comparing variables that can be divided

TIPS:

- Use contrasting colors
- Categorize data alphabetically, sequentially or by value

Profit from sales of items in different stores



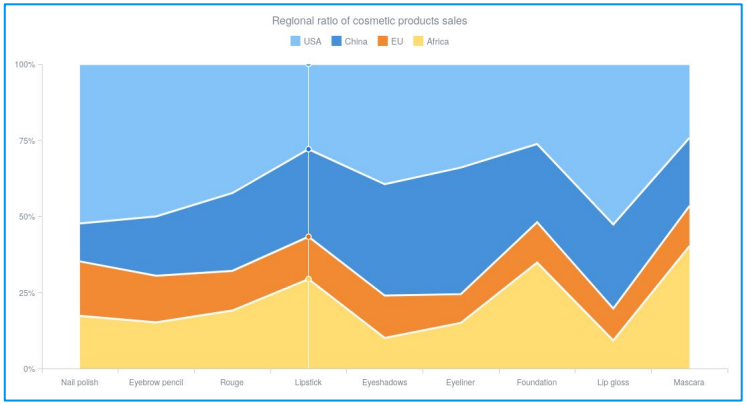
Stacked Area Chart

BEST FOR: Showing trends for several groups

TIPS:

- Use transparency to improve the chart's visibility

Regional ratio of cosmetic products sales



Distribution Plots

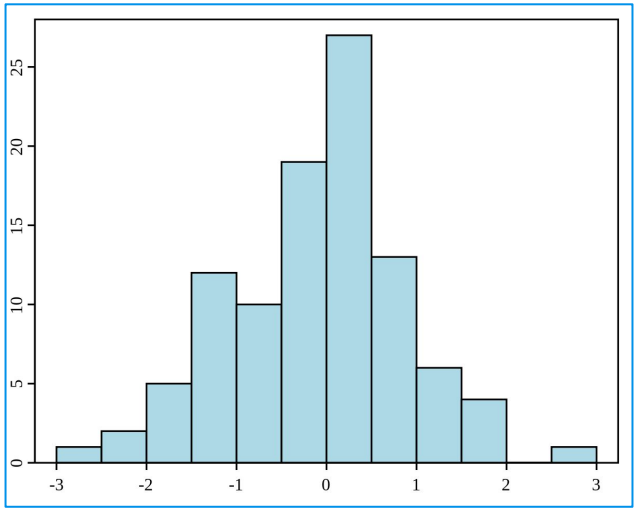
Used for visualizing the distribution of variables.

1. Histogram
2. Density Plot
3. Box Plot

Histogram

BEST FOR: Visualizing the distribution of a variable for a dataset

A simple histogram example



TIPS:

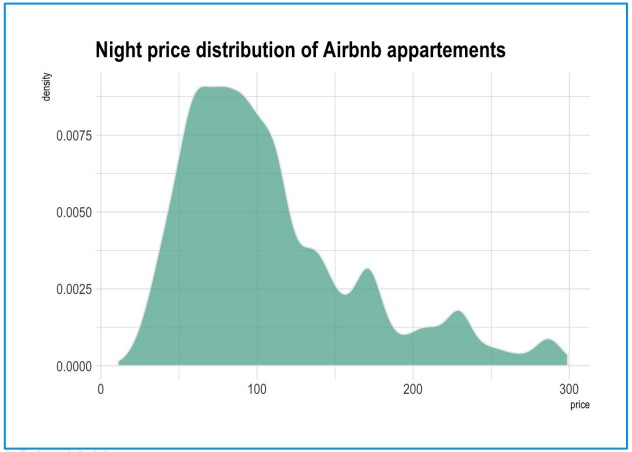
- Try different bin sizes (data intervals), the shape can vary significantly

DON'T CONFUSE IT WITH A BAR CHART

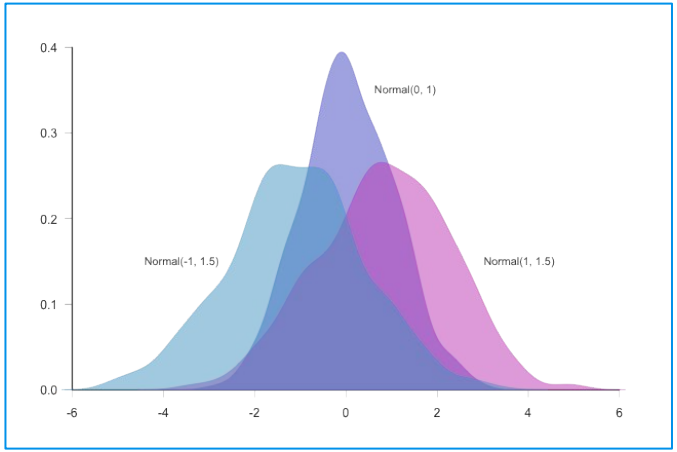
Density Plot

BEST FOR: Comparing the distribution of several variables. It is a smoothed variation of a histogram that doesn't depend on bin size.

Night price distribution of Airbnb appartements



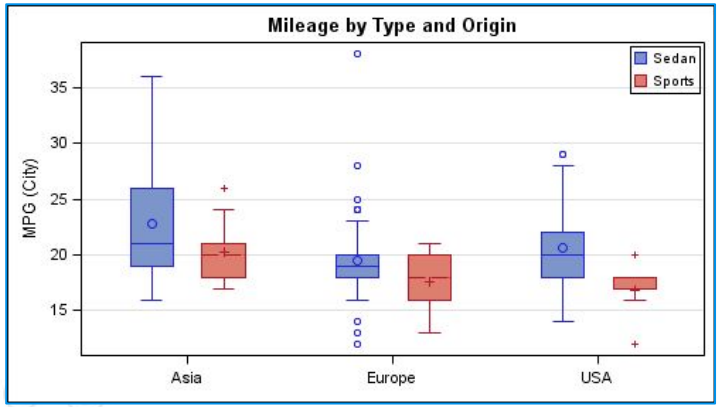
Example of density Plot

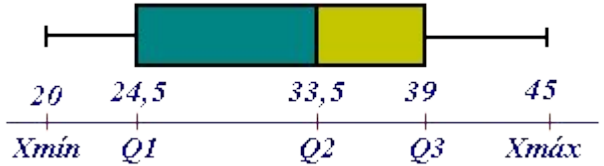


Box Plot

BEST FOR: Comparing statistical measures for multiple variables or groups. It represents the interquartile range (IQR), variability outside quartiles and, optionally, outliers.

Mileage by Type and Origin





- Whiskers:
  - Minimum to Q1
  - Maximum Q3
- **Q2** = Median
- **Points or diamonds** = Outliers

Geological Plots

Used to visualize geospatial data.

1. Dot Map
2. Choropleth Map
3. Connection Map

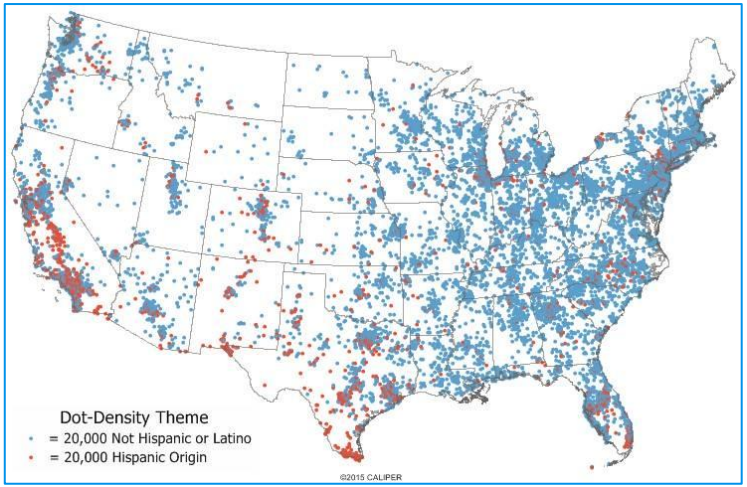
Dot Map

BEST FOR: Detecting spatial distributions in a geographic region. Every dot represents an observation and has the same size and value.

TIPS:

- The map should still be visible
- Adjust the dot size so that they blend in dense areas

Hispanic and not hispanic density



***Red* = Hispanic | *Blue* = Not Hispanic**

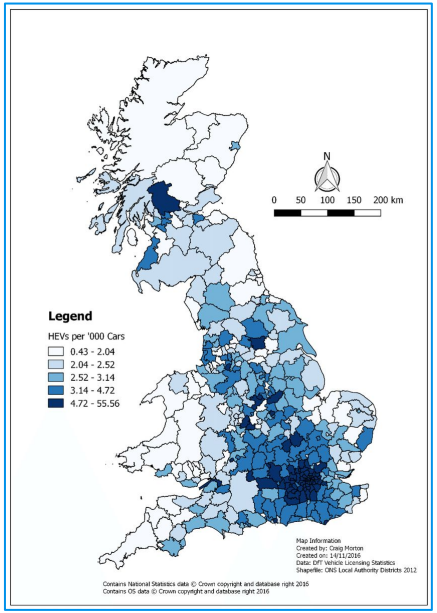
Choropleth Map

BEST FOR: Visualizing how a variable varies in a geographical area

TIPS:

- We pay attention to bigger areas → Normalize the data
- Use darker colors for higher values
- Don’t use too many colors

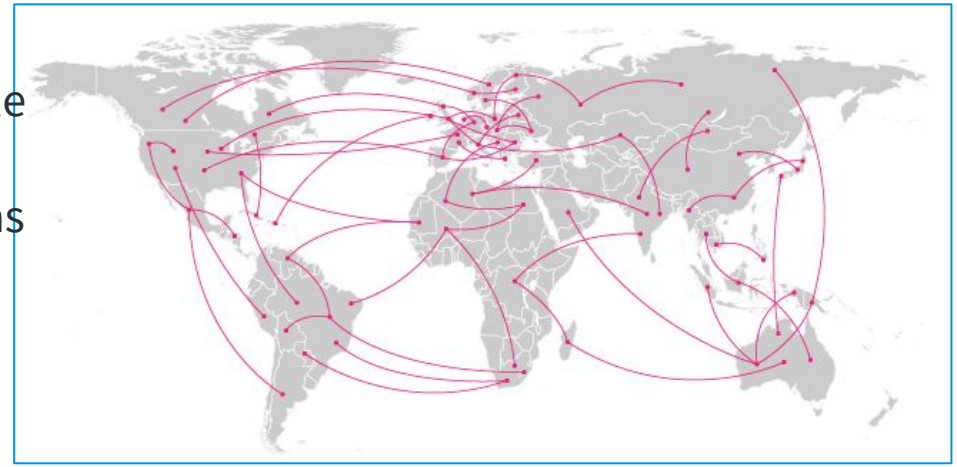
Hybrid Electric Vehicle Registrations



Connection Map

BEST FOR: Visualizing connections

Distances between different cities of the world



TIPS:

- The map should still be visible
- Adjust the line thickness so that they blend in dense areas

Summary

- Add a caption or a title
- Use labels for the axes
- Use colors to differentiate variables/categories
- Use color, shape and size to show additional variables
- Use labels for individual categorical variables

- **KEEP IT SIMPLE**, don’t add too much information

Bibliography

- The Data Visualization Workshop - Second Edition by Mario Dobler, Tim Großman
- Chart.js Documentation
- Data visualization with Chart.js: An introduction

Bibliography

- Create a line Chart in excel
- A complete guide to grouped bar Charts
- Scatterplot with line
- How to make bubble plot in python?
- 2D Color Pie Chart
- A complete guide to stacked bar charts

Bibliography

- A percent stacked area chart
- Wikipedia - Simple Histogram Example
- Density Plots
- What is a Dot-density Map
- Choropleth map illustrating Hybrid Electric Vehicle
- A Connection Map

Thanks

Any questions?

Contact us:

francisco.arocas.37@ull.edu.es

sara.revilla.13@ull.edu.es