Data Visualization

DataJam
Carmen Vaca, PhD. - Eduardo Cruz, MSc.
Dec 24th, 2018

1

Introduction

Are Visualizations important for Big Data?

DOMO

DATA NEVER SLEEPS

How Much Data Is Generated **Every Minute?**

Big data is not just some abstract concept used to inspire and mystify the IT crowd; it is the result of an avalanche of digital activity pulsating through cables and airwaves across the world. This data is being created every minute of the day through the most innocuous of online activity that many of us barely even notice. But with every website browsed, status shared, or photo uploaded, we leave digital trails that continually grow the hulking mass of big data. Below, we explore how much data is generated in one minute on the Internet.

THE MOBILE WEB RECEIVES 217 NEW USERS.

WORDPRESS USERS PUBLISH

YOUTUBE USERS UPLOAD 48 HOURS OF NEW VIDEO.

EMAIL USERS SEND 204,166,667 MESSAGES.

GOOGLE RECEIVES OVER 2,000,000 SEARCH QUERIES.

FACEBOOK USERS

David McCandless,

British information designer.

TEDGlobal, 2010

https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization

4

> *All of us now are being blasted by information design. It's being poured into our eyes through the Web, and we're all visualizers now; we're all demanding a visual aspect to our information… And if you're navigating a dense information jungle, coming across a beautiful graphic or a lovely data visualization, it's a relief, it's like coming across a clearing in the jungle.*
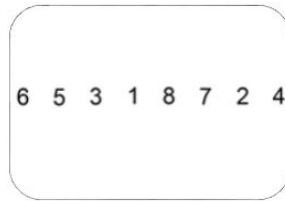
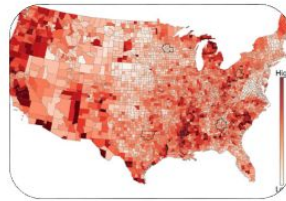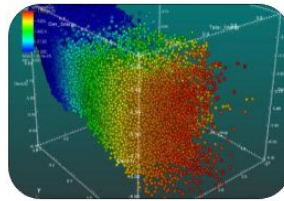taws
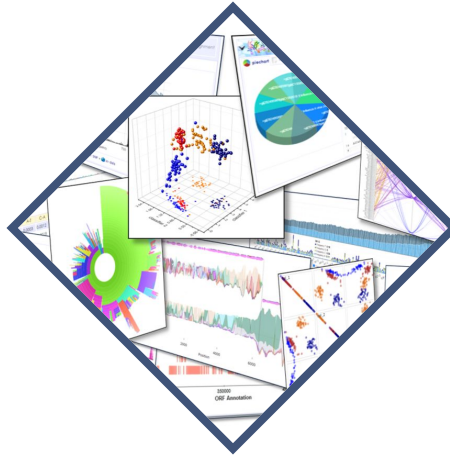
Visualizations?

YES!

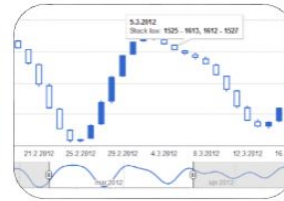# 2

General Concepts

Types and Uses

Types

Linear
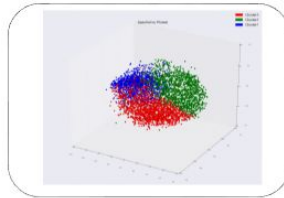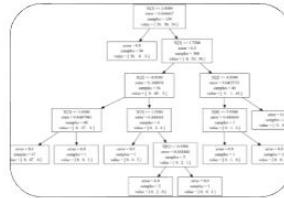

Planar (GeoSpatial)

Volumetric

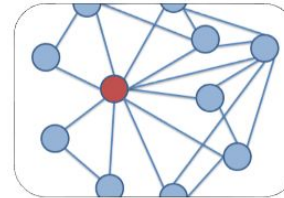
Temporal

nD/Multidimensional

Tree - Hierarchical

Network

https://guides.library.duke.edu/datavis/vis_types

Planar

Choroplet. Cartogram. Dot distribution map. Proportional symbol map. More…

3

GeoVisualizations

Applications

Choroplet

Maps

- Healthcare: prevention and mapping.
- **Socieconomic indices.**
- Any quantity measure.

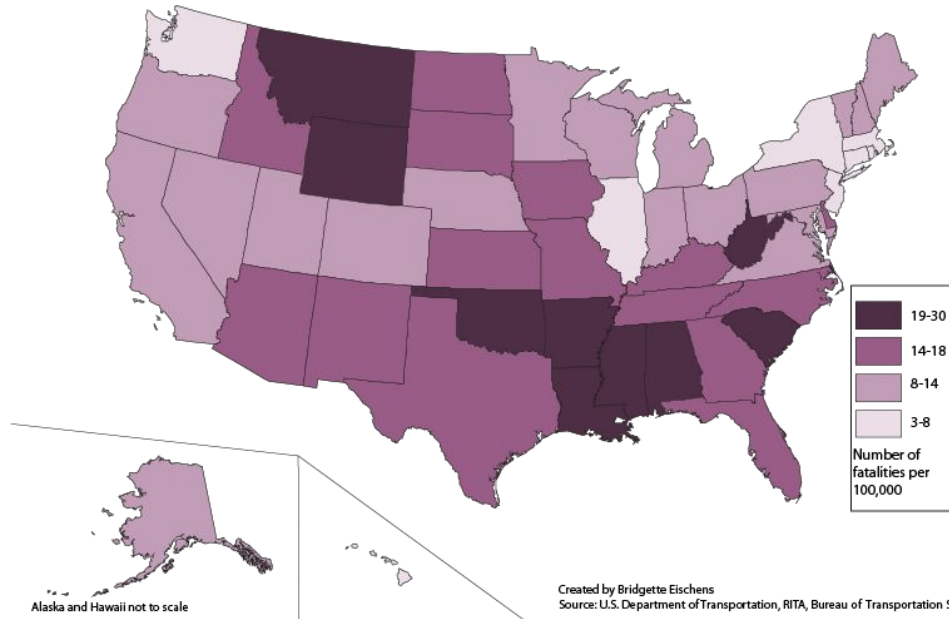U.S. Motor Vehicle Fatalities, 2008
choropleth map using standard deviation classification

19-30
14-18
8-14
3-8

Number of fatalities per 100,000

Alaska and Hawaii not to scale

Created by Bridgette Eischens
Source: U.S. Department of Transportation, RITA, Bureau of Transportation Statistics

The use of geovisualization to public health, in the context of open source applications and digital earths: an effective representation?

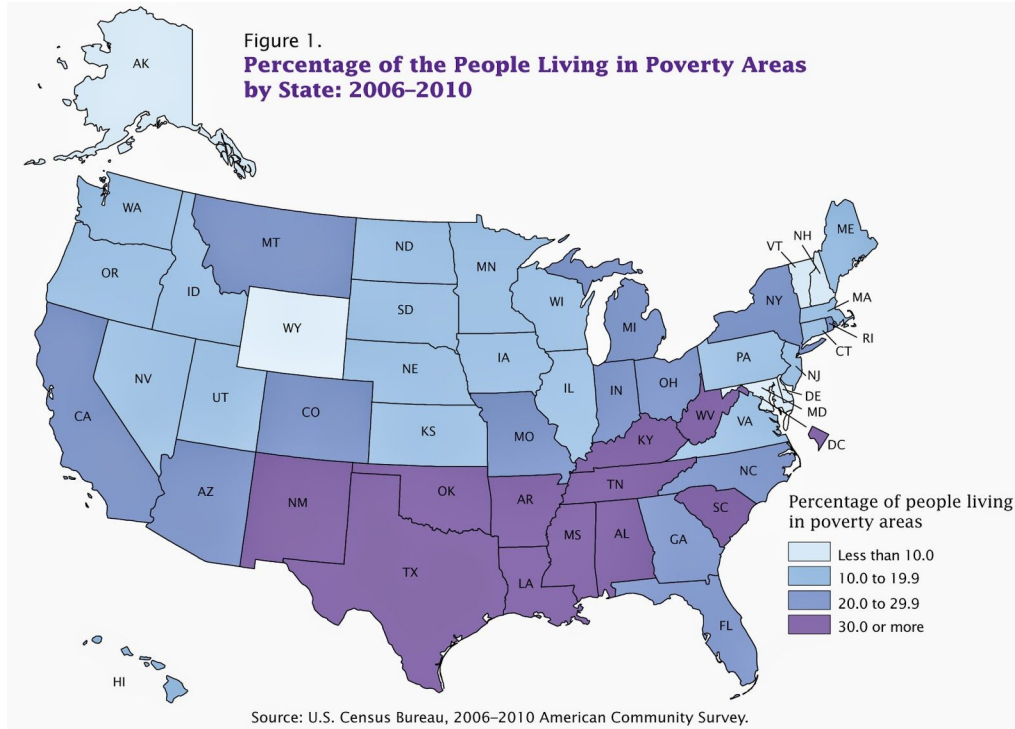Ana Brandusescu, Renee Sieber, Nadine Schuurman

SOCIOECONOMIC INDICES

Figure 1.
Percentage of the People Living in Poverty Areas by State: 2006–2010

Percentage of people living in poverty areas

- Less than 10.0
- 10.0 to 19.9
- 20.0 to 29.9
- 30.0 or more

Source: U.S. Census Bureau, 2006–2010 American Community Survey.

4

Bokeh as Interactive Visualizations Tool

Crime Case Study

Bokeh

Examples

Get ready!

Installation

pip install bokeh

Example Code

[Go to GitHub!](#)

Bokeh

Basic Elements
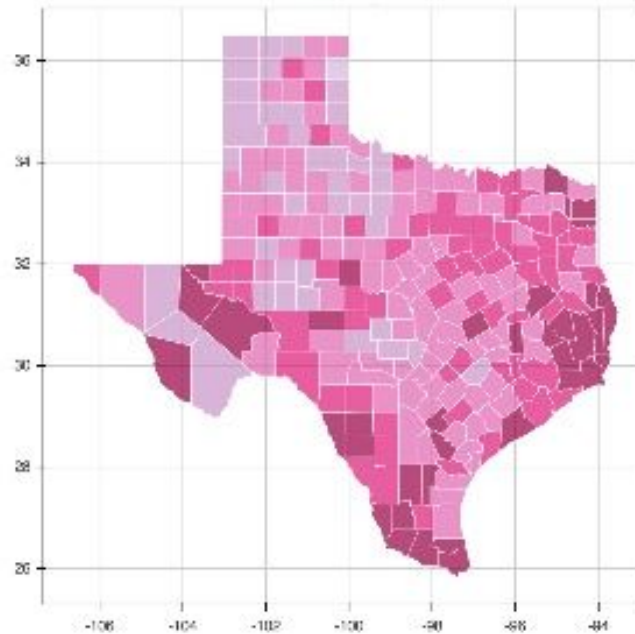
- bokeh.models

- bokeh.plotting

- bokeh.palletes

The basic element in Bokeh is a *figure* como en pandas es un Dataframe.

Geovisualization

Temporal Visualizations

For example:

Time #1, observation
Time #2, observation
Time #3, observation

1

Useful Concepts

Components of Time Series

- **Level.** The baseline value for the series if it were a straight line

- **Trend.** The optional and often linear increasing or decreasing behavior of the series over time

- **Seasonality.** The optional repeating patterns or cycles of behavior over time

- **Noise.** The optional variability in the observations that cannot be explained by the model.

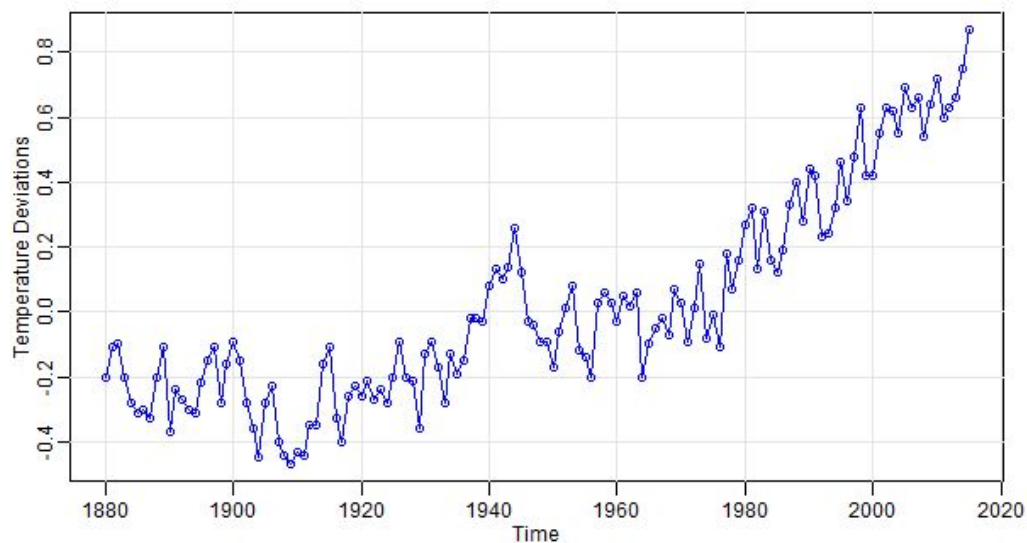**Time series data often requires cleaning, scaling, and even transformation.**

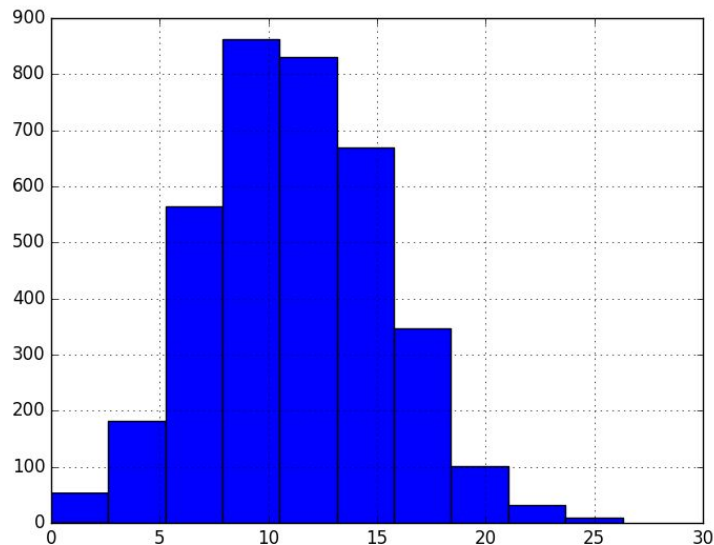- **Frequency.**

- **Outliers.**

- **Missing.**
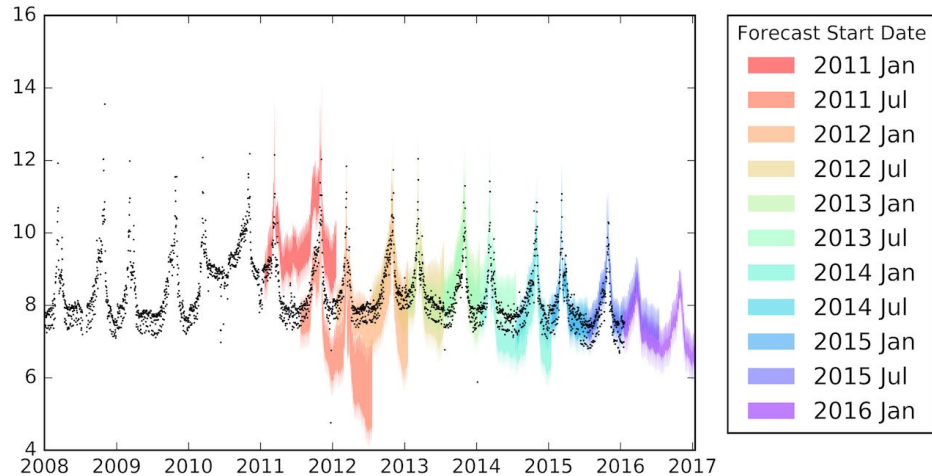
- **1. Line Plots.**

- **2. Histograms and Density Plots.**

3. Box and Whisker Plots.
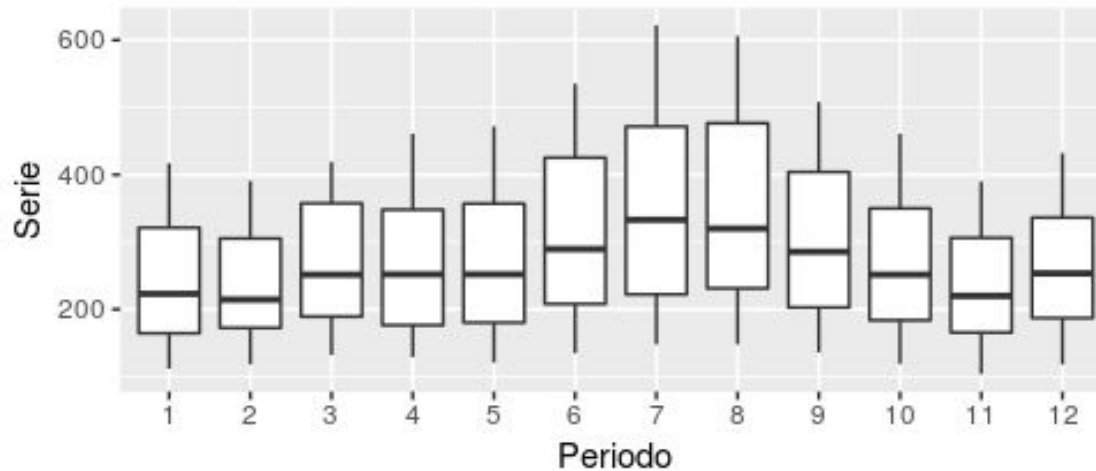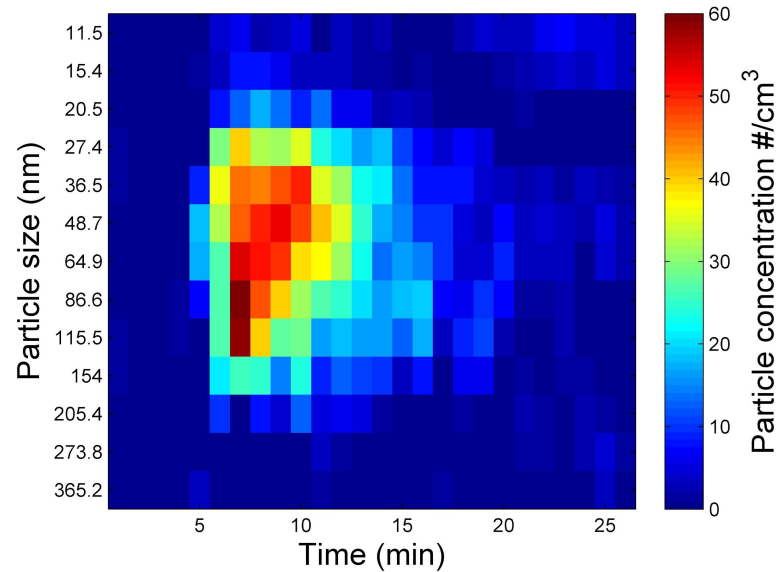
4. Heat Maps.

5. Lag Plots or Scatter Plots.
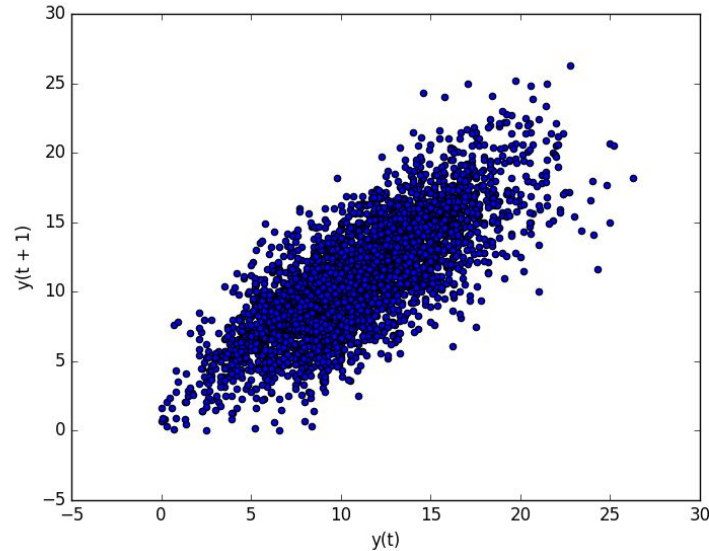
6. Autocorrelation Plots

Partial Autocorrelation
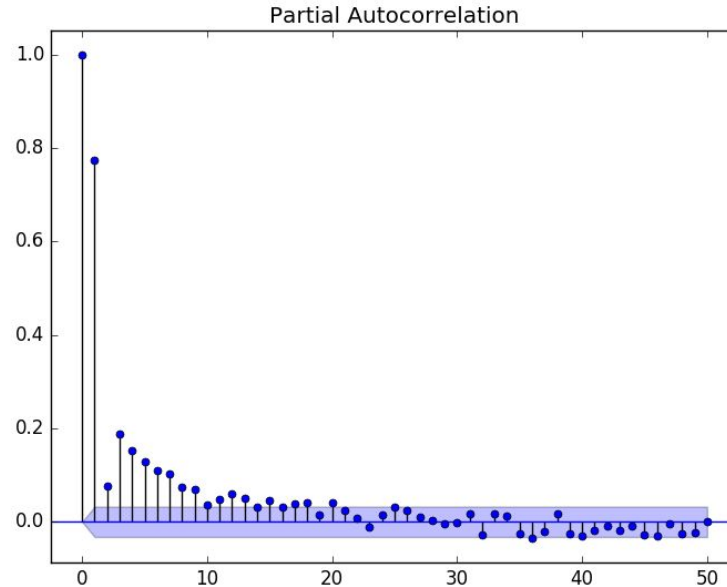
Get ready!

Installation

- Pandas
- Matplotlib

Example Code

[Go to GitHub!](#)

THANKS!

Any questions?