upGrad

#LifeKoKaroLift

PG Diploma in ML

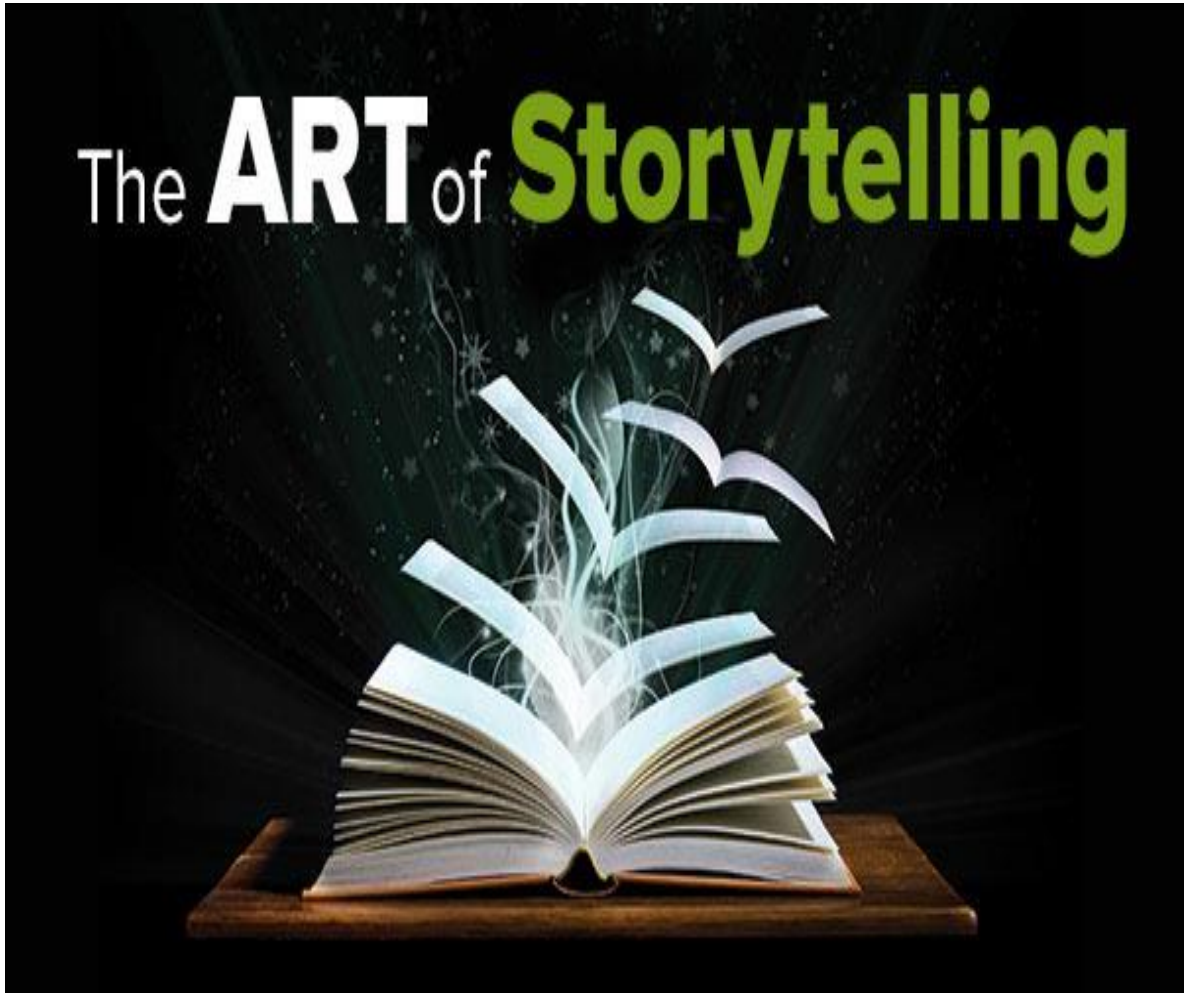
upGrad

Course : PG Diploma in ML

Lecture On : Live Session on Data Story Telling

Instructor : Arihant Jain

upGrad



Instructor-led live session on data storytelling

upGrad

About Me

I am working as **Lead Data Scientist** at ZestMoney with Close to 6+ years of experience in Data Science, Deep Learning, predictive Modelling, Machine Learning & Optimization. I worked for building Machine Learning / Deep Learning models in Retail, Credit Risk ,Marketing ,Collection ,Customer Service & Digital across Banking ,Digital, Industrial IOT & Telecom domain. I have worked with companies Genpact , RBL Bank , GE & Vodafone in Past , using Data Science & Analytics to drive different impact business use cases.

I am passionate about solving business problems through data science**. I believe every number has a story to tell. Being a data scientist it's my job and passion to decode that story.** I enjoy applying and optimizing classical machine learning algorithms, NLP, and Bayesian design strategy to solve real-world problems.

Arihant Jain

https://www.linkedin.com/in/arihantjain15/

upGrad

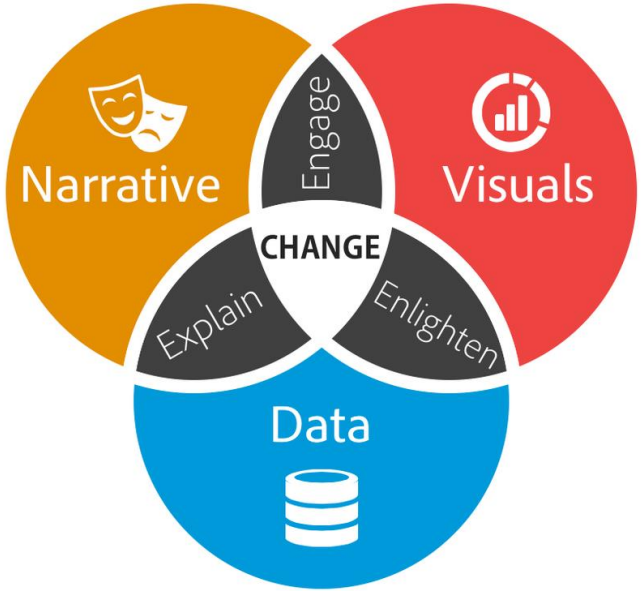
Agenda

- What is Storytelling ?
- Why Storytelling is essential
- Different Methods to create stories
- Types of dataset & appropriate charts
- Storytelling during Predictive Model building
- Best Practices of Storytelling
- Demonstration
- Q & A

What is Data Storytelling?

Data visualization expert **Stephen Few** said, “Numbers have an important story to tell. They rely on you to give them a clear and convincing voice”

Data storytelling is a structured approach for communicating data insights, and it involves a combination of three key elements: ***data***, ***visuals***, and ***narrative***.

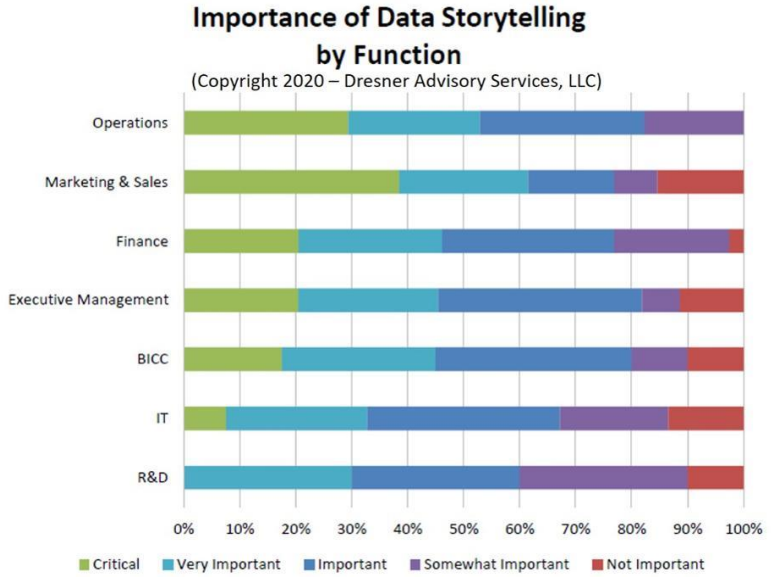


Why Storytelling is essential ?

Memorability- A study by Stanford professor Chip Heath found 63% could remember stories, but only 5% could remember a single statistic. *People hear statistics, but they feel stories*

Persuasiveness

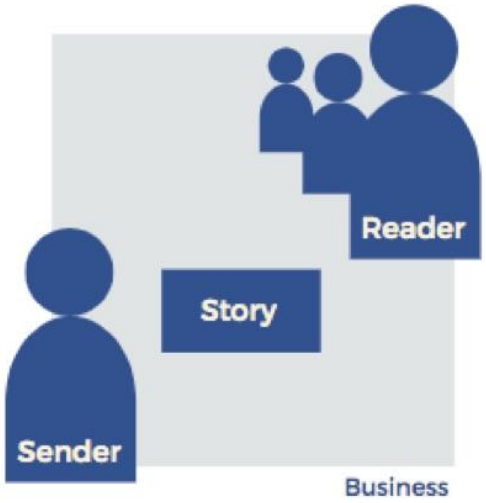
Engagement



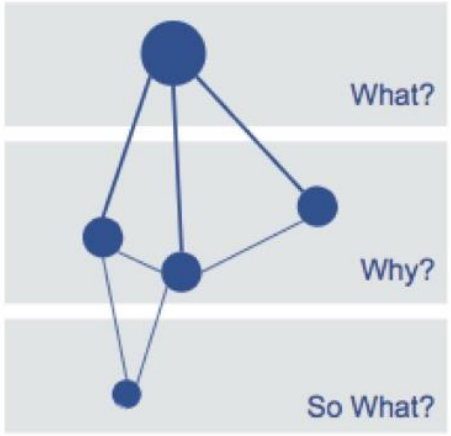
upGrad

Flow of Storytelling

Context

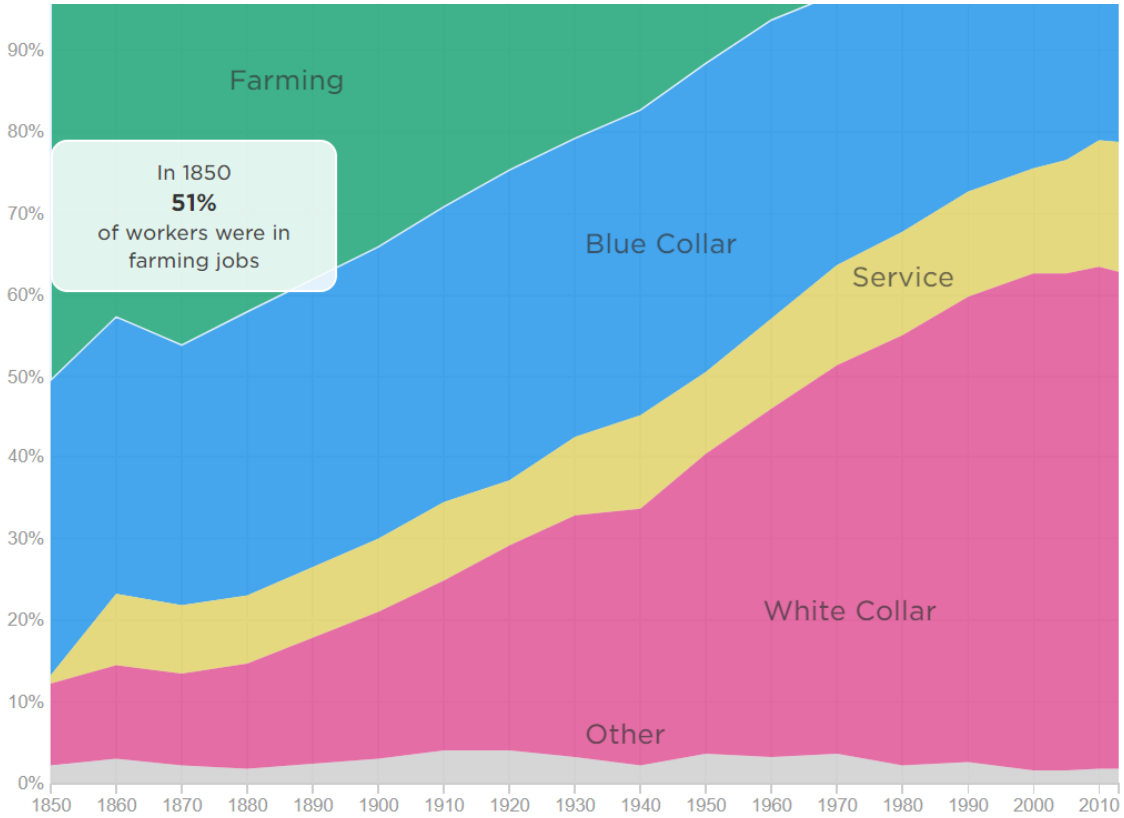


Narrative



Effective Visuals

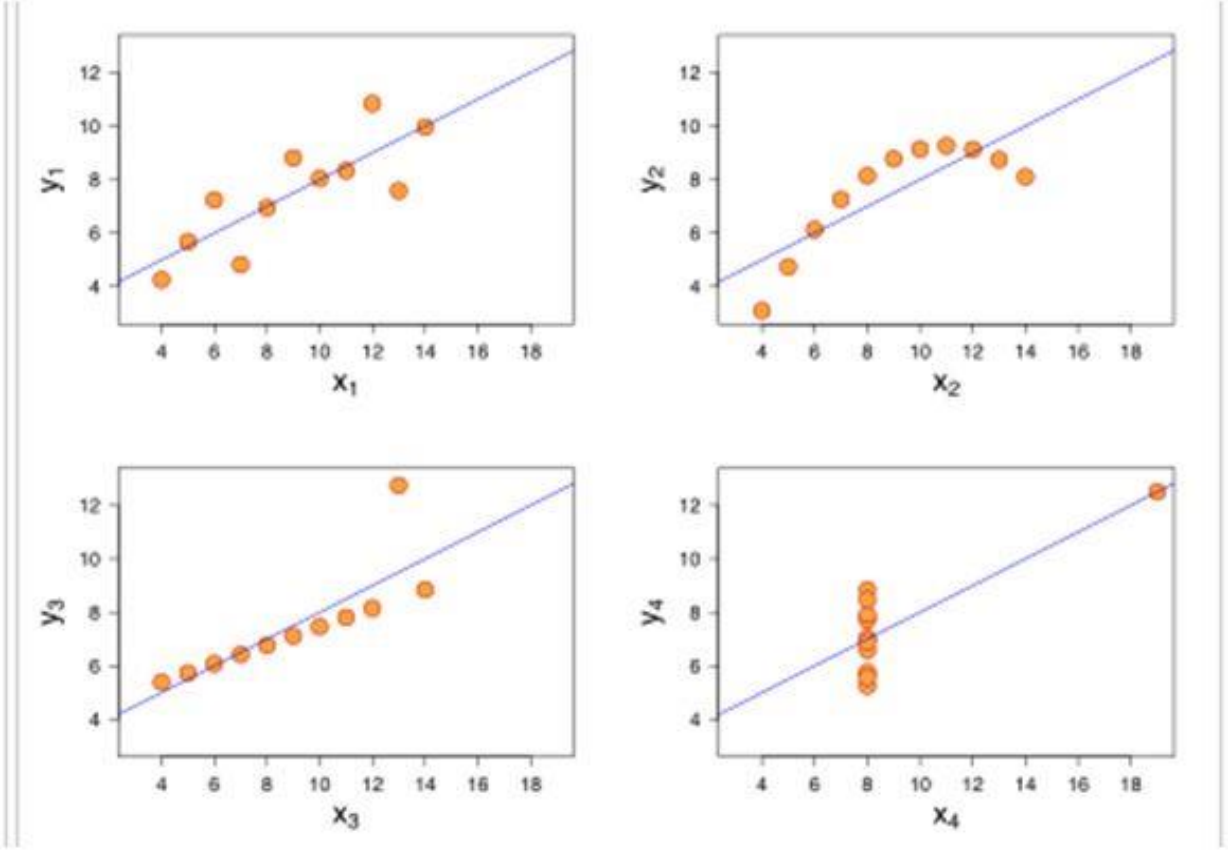
upGrad



How Machines Destroy (And Create!) Jobs

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89


y1
x1
y2
x2
y3
x3
y4
x4
4
6
8
10
12
14
16
18

How Seattle Residents determine housing. In one simple graph, they clearly communicate three variables: neighborhood, amenity, and scale.

GRAPH LEGEND

- Remainder of Seattle
- Eastside
- South King County
- Snohomish
- Central Seattle

NOTE*

Walkability: Walkable neighorhood and being near local activities
Space and separation: Having space and separation from others
Commute: Within 30-minute commute to work
Change in family size: Change in family size or marital/partner status

Source: Puget Sound Regional Council

Step Wise Plan for Story Telling

upGrad

1. Begin with a question
2. End with the Insight
3. Tell Smart Story
4. Use Visuals to support
5. Know your Audience and cater your story accordingly

upGrad

Different Methods to create stories

1.Begin with Pen-Paper Approach

- Scripting down your ideas and flow before you start structuring your story is very essential to your final product

Aristotle's classic five-point plan that helps deliver strong impacts is:

- Deliver a story or statement that arouses the audience's interest.
- Pose a problem or question that has to be solved or answered.
- Offer a solution to the problem you raised.
- Describe specific benefits for adopting the course of action set forth in your solution.
- State a call to action.

2.Dig deeper to identify the sole purpose of your story

- Identify closely, what the idea of your story is. Ask yourself, “What am I really giving with this story?” It’s never the story alone, but what the story can do to make decision making better. What you’re displaying is the idea of a better decision making or analytics.

3.Design a Road Map

- Create a list of all the key points you want your audience to know about your story, visual, or analysis.
- Categorize the list until you are left with only three major message points. This group of three will provide the verbal road map for your story.
- Under each of your three key messages, add supporting evidence to enhance the narrative. These could include some or all of the following: personal stories, facts, examples, analogies etc.

Types of Data and Suitable Charts

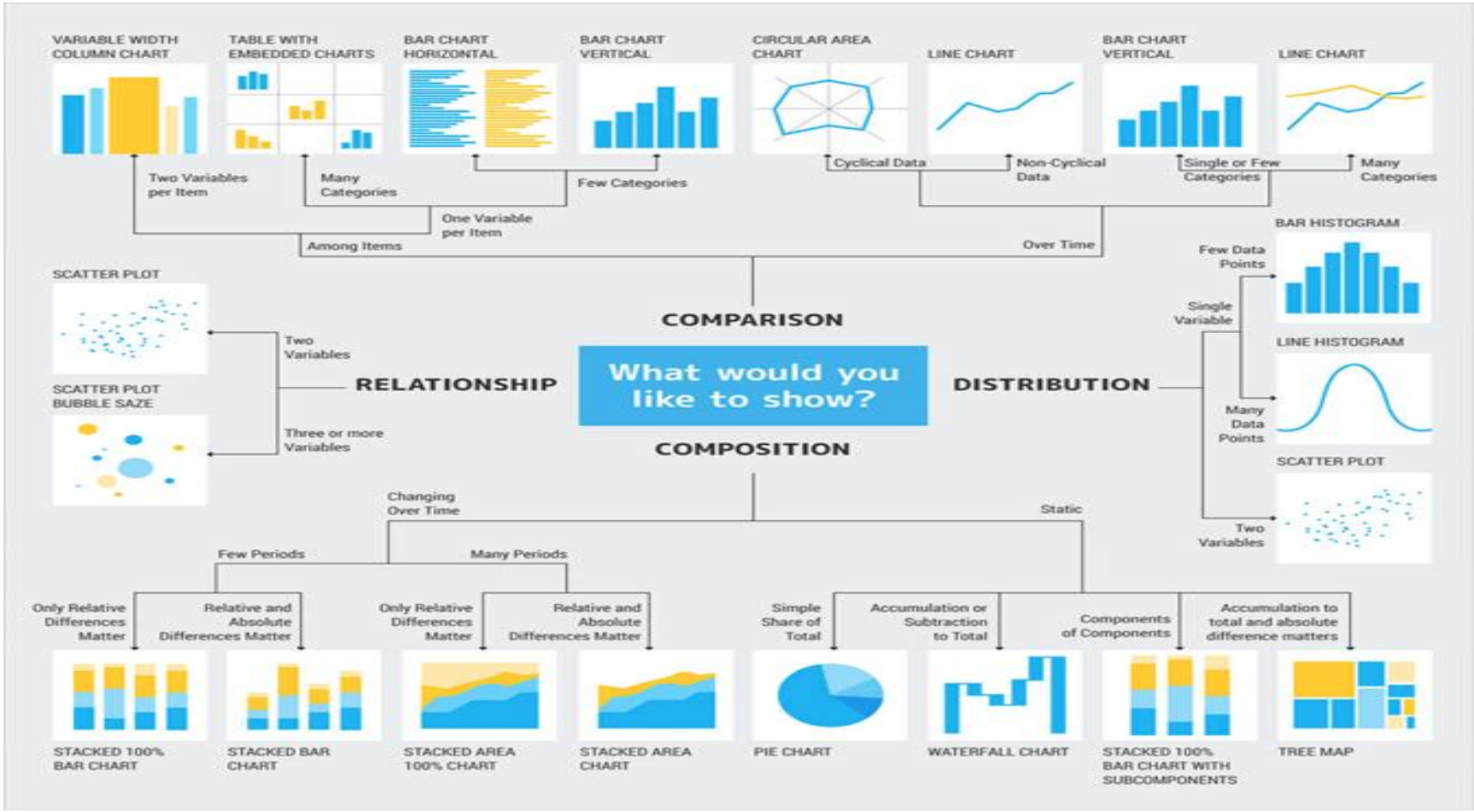
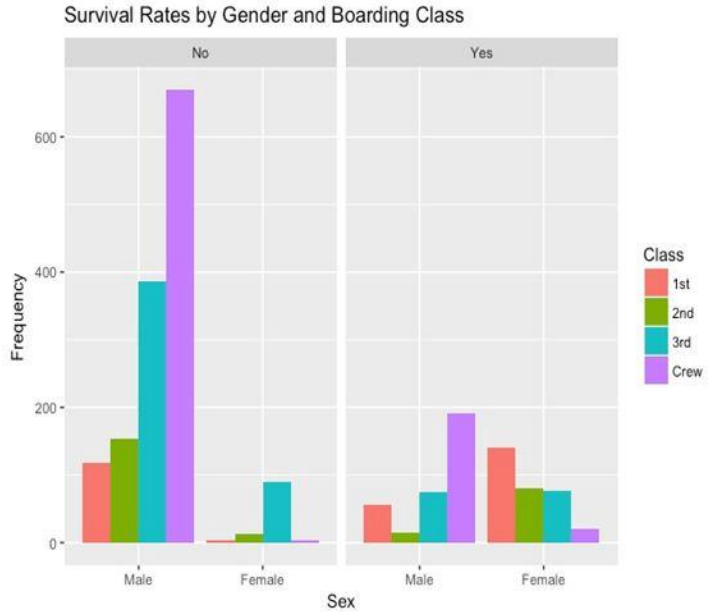


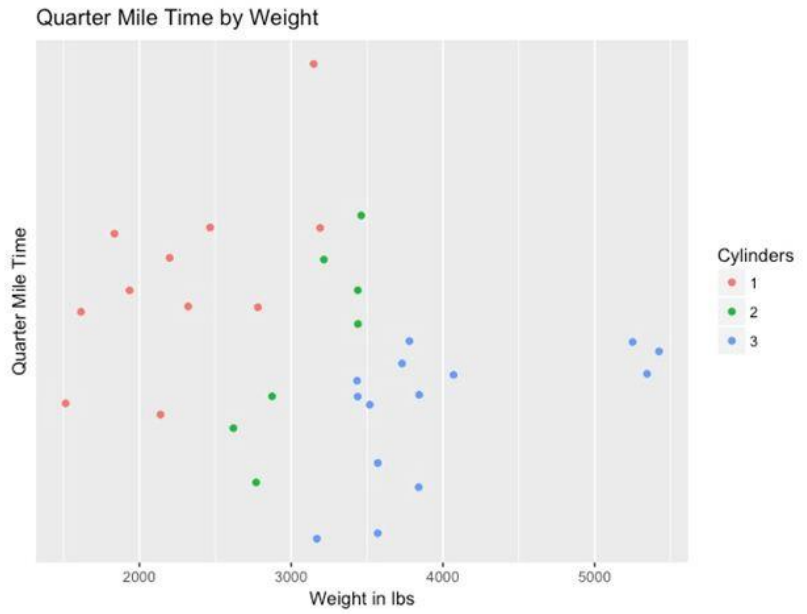
Image Source – Crazy Egg

1. Textual Data

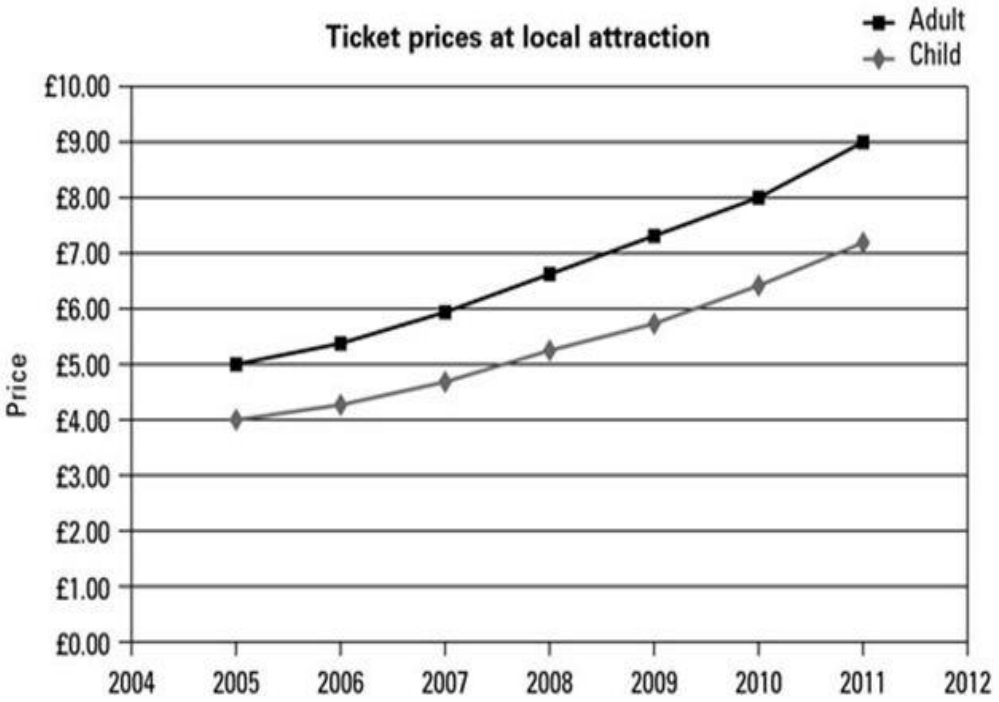
- One of the best-suited visualizations for textual data is the WordCloud. The wordcloud brings the more frequent ones to the center and enlarges them, giving us a clear picture of what the general idea of the text depicts.

2. Dataset consist of Continuous & Categorial Data





3. Numerical Dataset



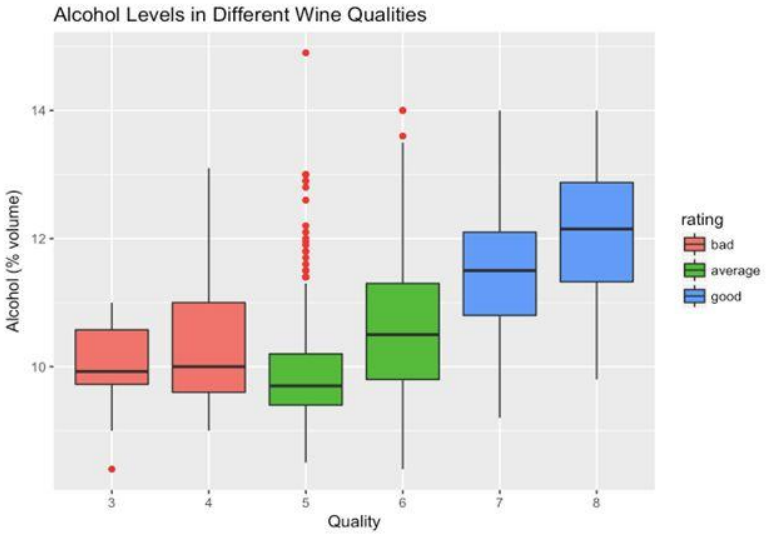
4. Time Series Dataset

5. Geospatial Dataset

Storytelling during the steps of predictive modeling

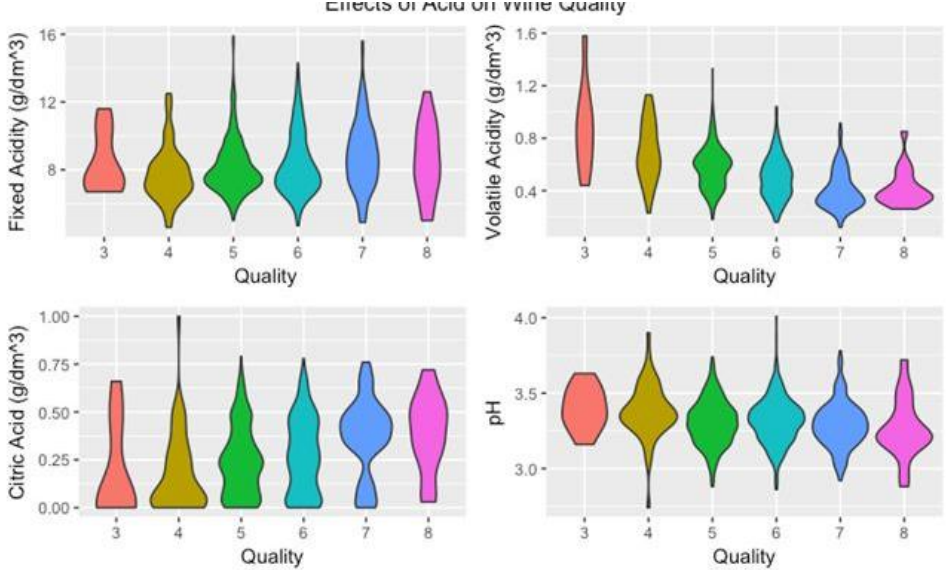
1. Data Exploration

Let’s consider a dataset on Wine Quality

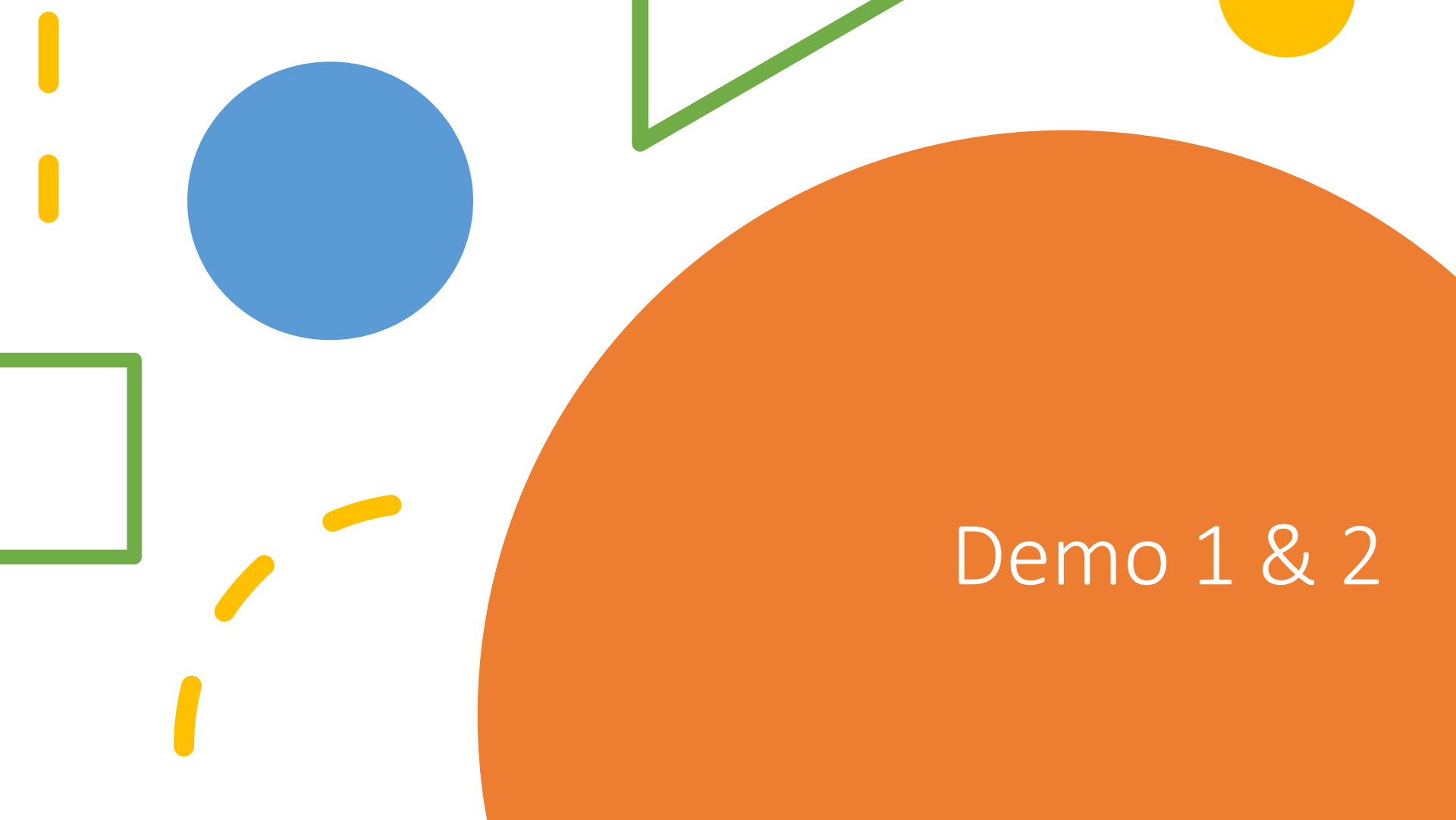



```
bs. of  15 variables:
 : int  1 2 3 4 5 6 7 8 9 10 ...
 : num  7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 7.5 ...
 : num  0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.5 ...
 : num  0 0 0.04 0.56 0 0 0.06 0 0.02 0.36 ...
 : num  1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 6.1 ...
 : num  0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065 0.073 0.071
 : num  11 25 15 17 11 13 15 15 9 17 ...
e: num  34 67 54 60 34 40 59 21 18 102 ...
 : num  0.998 0.997 0.997 0.998 0.998 ...
 : num  3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.35 ...
 : num  0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.8 ...
 : num  9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 10.5 ...
 : Ord.factor w/ 6 levels "3"<"4"<"5"<"6"<..: 3 3 3 4 3 3 3 5 5 3 .
 : Ord.factor w/ 3 levels "bad"<"average"<..: 2 2 2 2 2 2 2 3 3 2 .
 : num  8.1 8.68 8.6 12.04 8.1 ...
```

Data Exploration



- Next, would you wonder how acid contents in your wine affect its quality?


Demo 1 & 2

Best Practices for Story Telling

- Always **label your axes and give the heading** of your plot.
- Use **legends** where necessary.
- Use **colors that are lighter** on the eye and in proportion.
- **Avoid adding unnecessary detail** to your visualization like backgrounds or themes that don't allow good readability.
- **Only a point can be used** to simultaneously encode two quantitative values based on a horizontal and
- **Never use points** for visualization if you are doing time series encoding.

upGrad

#LifeKoKaroLift

Thank You!