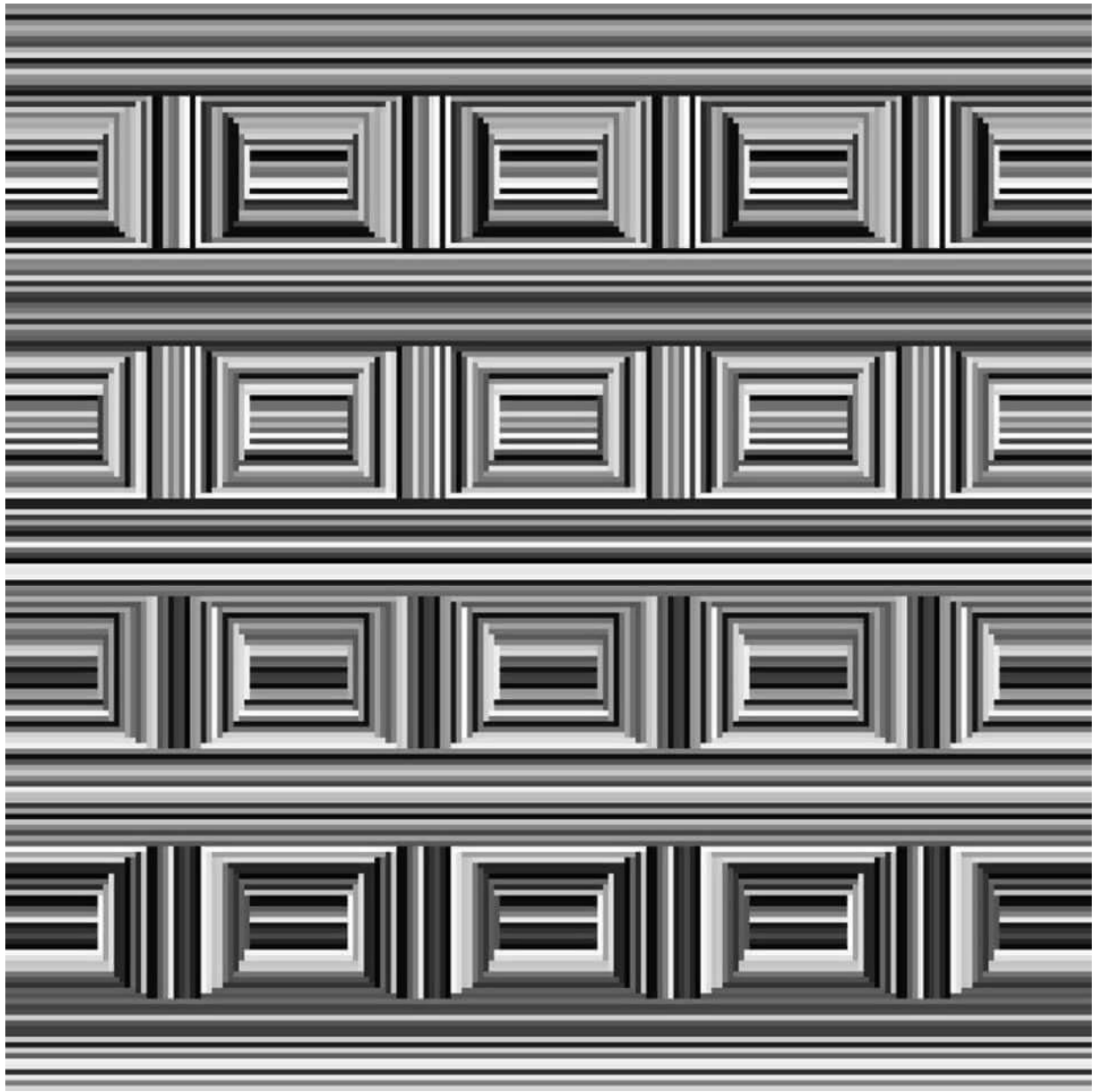
Data Visualization



What is Data Visualization?

- Data (or information) visualization is used to interpret and gain insight into large amounts of data. This is achieved through visual representations, often interactive, of raw data.
- It is specific technique or methodology for creating pictorial representations in the form of images, diagrams or animations to convey effective information from data.

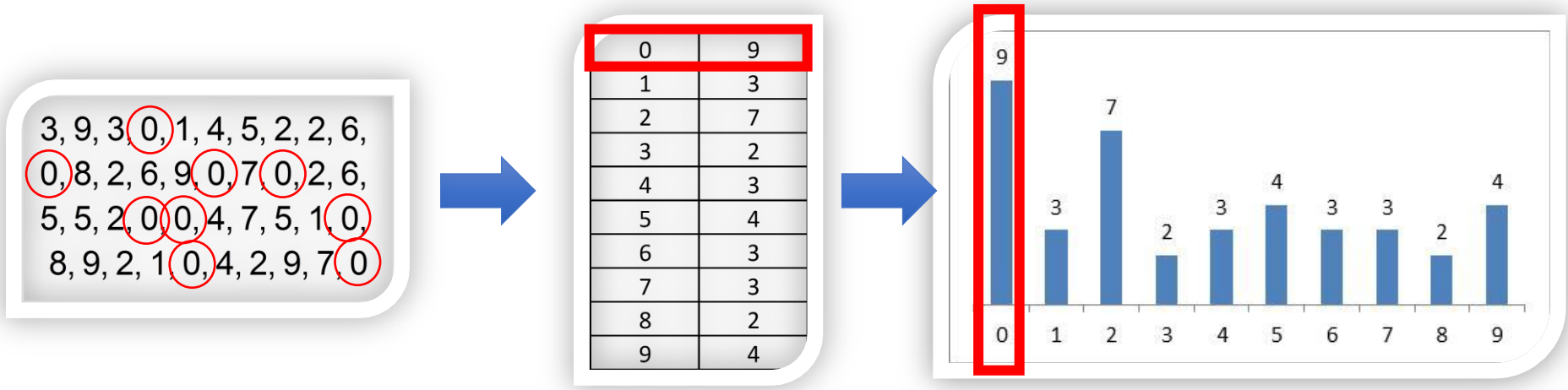
Example

How many 0's in the data?

0,2,3,0,6,0

From this data we can easily say that there **3** 0's in the data but when we have huge data like below

3, 9, 3, 0, 1, 4, 5, 2, 2, 6, 0, 8, 2, 6, 9, 0, 7, 0, 2, 6, 5, 5, 2, 0, 0, 4, 7, 5, 1, 0, 8, 9, 2, 1, 0, 4, 2, 9, 7, 0



Why is Data Visualization?

90% of information transmitted to the human brain is visual

The human brain processes visuals 60K times faster than text

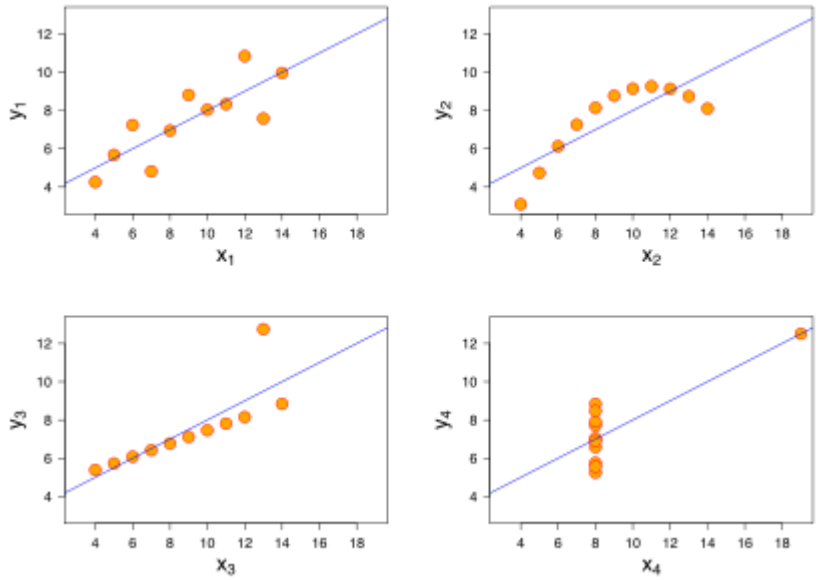
65% of humans are visual learners

The human brain can process an observed visual in 13 - 80 ms

Why is Data Visualization?

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.1	10	7.46	8	6.6
8	6.95	8	8.1	8	6.77	8	5.8
13	7.58	13	8.7	13	12.7	8	7.7
9	8.81	9	8.8	9	7.11	8	8.8
11	8.33	11	9.3	11	7.81	8	8.5
14	9.96	14	8.1	14	8.84	8	7
6	7.24	6	6.1	6	6.08	8	5.3
4	4.26	4	3.1	4	5.39	19	13
12	10.8	12	9.1	12	8.15	8	5.6
7	4.82	7	7.3	7	6.42	8	7.9
5	5.68	5	4.7	5	5.73	8	6.9

- Mean of x in each case: 9 (exact)
- Sample variance of x in each case: 11 (exact)
- Mean of y in each case: 7.50 (to 2 decimal places)
- Sample variance of y in each case: 4.122 or 4.127 (to 3 decimal places)
- Correlation between x and y in each case: 0.816 (to 3 decimal places)
- Linear regression line in each case: y = 3.00 + 0.500x (to 2 and 3 decimal places, respectively)



Data Visualization:

- **Enhances learning**
- **Enhances understanding**
- **Enhances reasoning**
- **Helps in decision making**

Data visualization acts as a link between the raw data and our engagement with it.

Data Visualization..

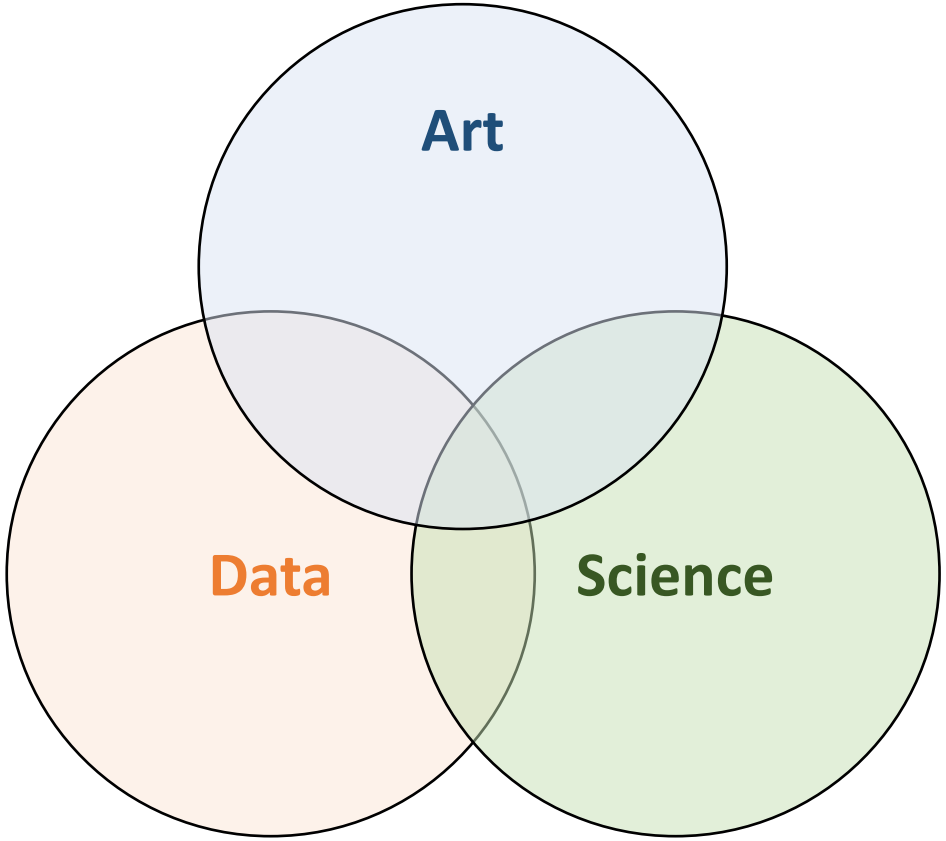
Can lead a user to

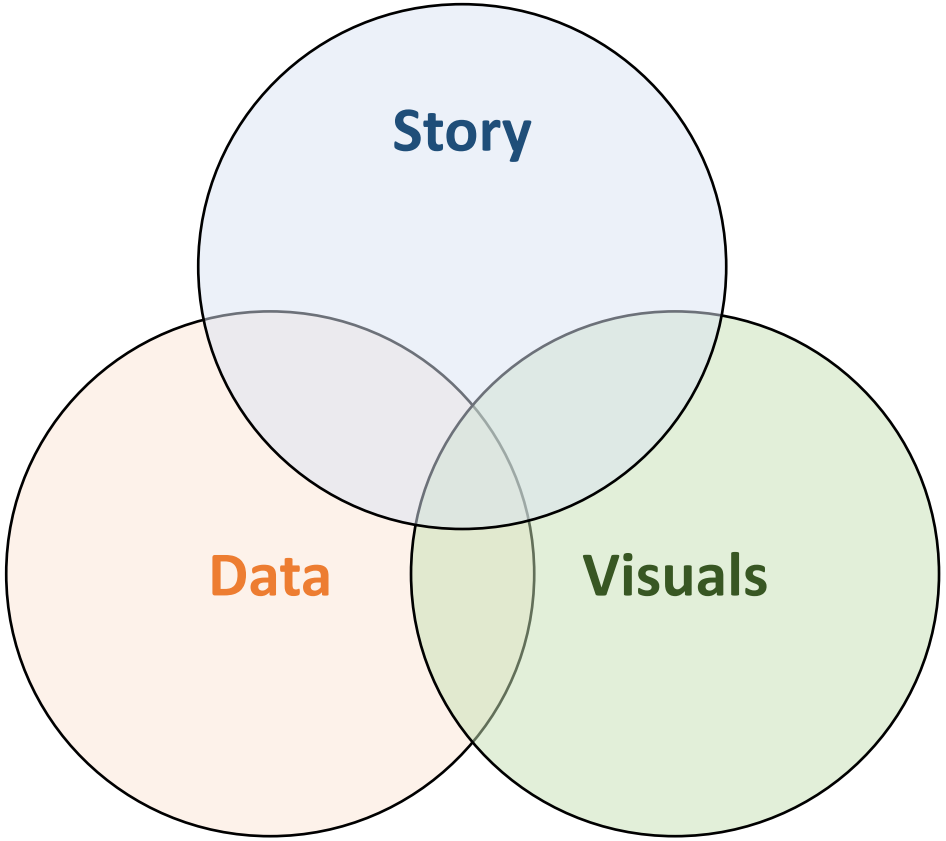
- Detect patterns
- Detect trends
- Detect correlations in data

Can then prompt a user to

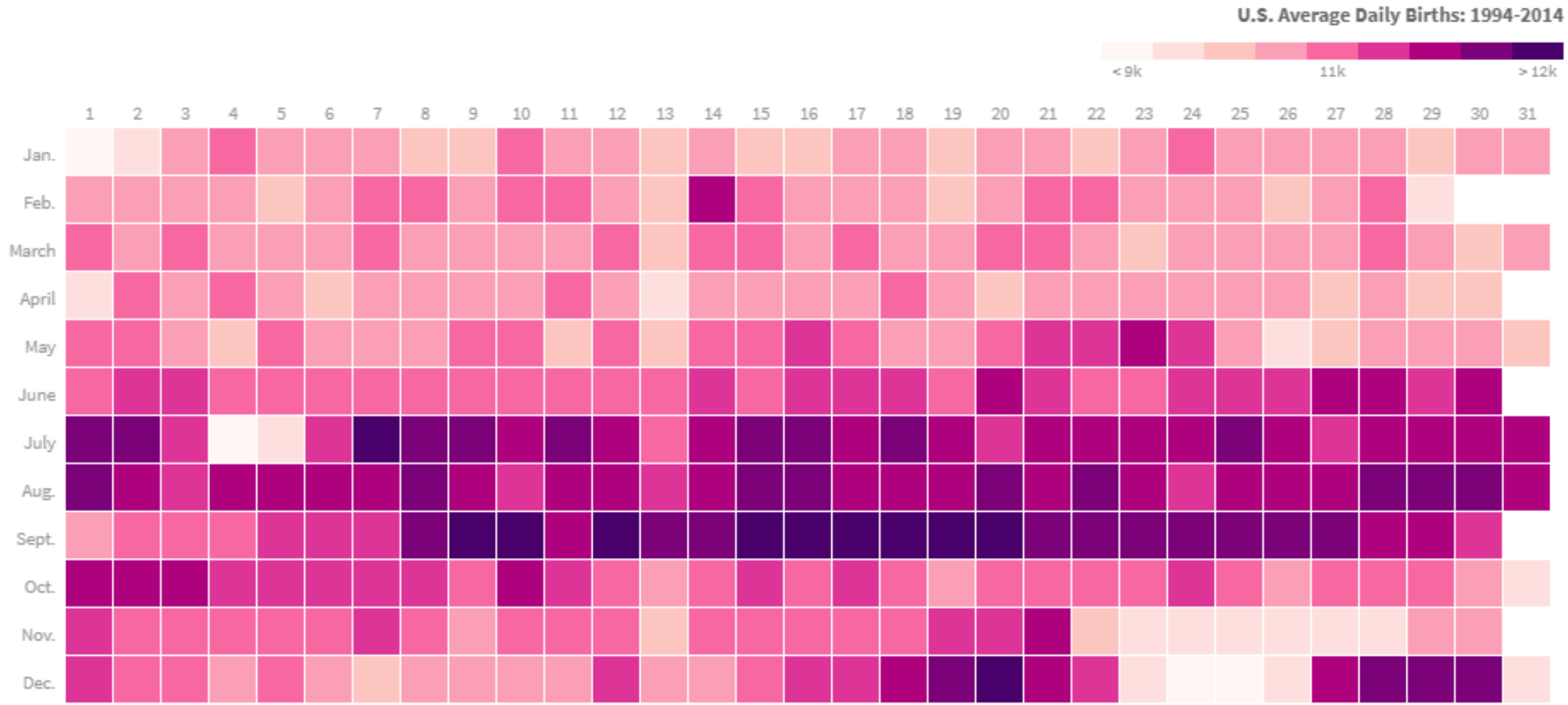
- Draw inferences
- Anticipate potential trajectories and outcomes
- Ask new questions of the data that wouldn't have otherwise been considered

Data Visualization & Storytelling





Example


U.S. Average Daily Births: 1994-2014
< 9k
11k
> 12k
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Jan.
Feb.
March
April
May
June
July
Aug.
Sept.
Oct.
Nov.
Dec.

Some Highlights:

- **Holidays:** People generally seem to have time for baby making during their time off. Several of the most-common birth dates, in September, correspond with average conception periods around Christmas. Sept. 9 is most-common in this dataset, though other days in that month are close. Sept. 19 is second. Following a customary gestation period, many of these babies would, in theory, have been conceived on Dec. 17 and 27, respectively.

- **Choice:** Clearly, some people choose when they have their children. While they're making babies during the holidays, many people aren't really having them then. The least-common birthdays in this dataset were Christmas Eve, Christmas and New Year's Day. Dates around Thanksgiving aren't as common. July 4 is also at the bottom of the list. Conversely, Valentine's Day ranks relatively high, as you can see in the graphic, as are the days just before a new tax year begins.

- **Skewed:** There are some fun patterns in this data, but the difference between birthday — unless it's on a truly rare day — isn't *that* much different than a top date in September. There a left-tailed skewness to the data, which ranges from 6,500 births per day to more than 12,000. The median number of births per day, though, is around 11,000. The most-common day had 12,300 births, on average. *More on the data distribution soon*.

Contd..

BIRTH DATE	POPULARITY	CONCEPTION*	AVERAGE BIRTHS
9/9	1st	12/17	12,301
9/19	2nd	12/27	12,229
9/12	3rd	12/20	12,224
9/17	4th	12/25	12,148
9/10	5th	12/18	12,143
7/7	6th	10/14	12,108
9/20	7th	12/28	12,107
9/15	8th	12/23	12,087
9/16	9th	12/24	12,072
9/18	10th	12/26	12,055
12/20	11th	3/29	12,009
9/26	12th	1/3	11,993
9/8	13th	12/16	11,992
9/23	14th	12/31	11,974
12/29	15th	4/7	11,956
8/8	16th	11/15	11,951
9/24	17th	1/1	11,945

Contd..

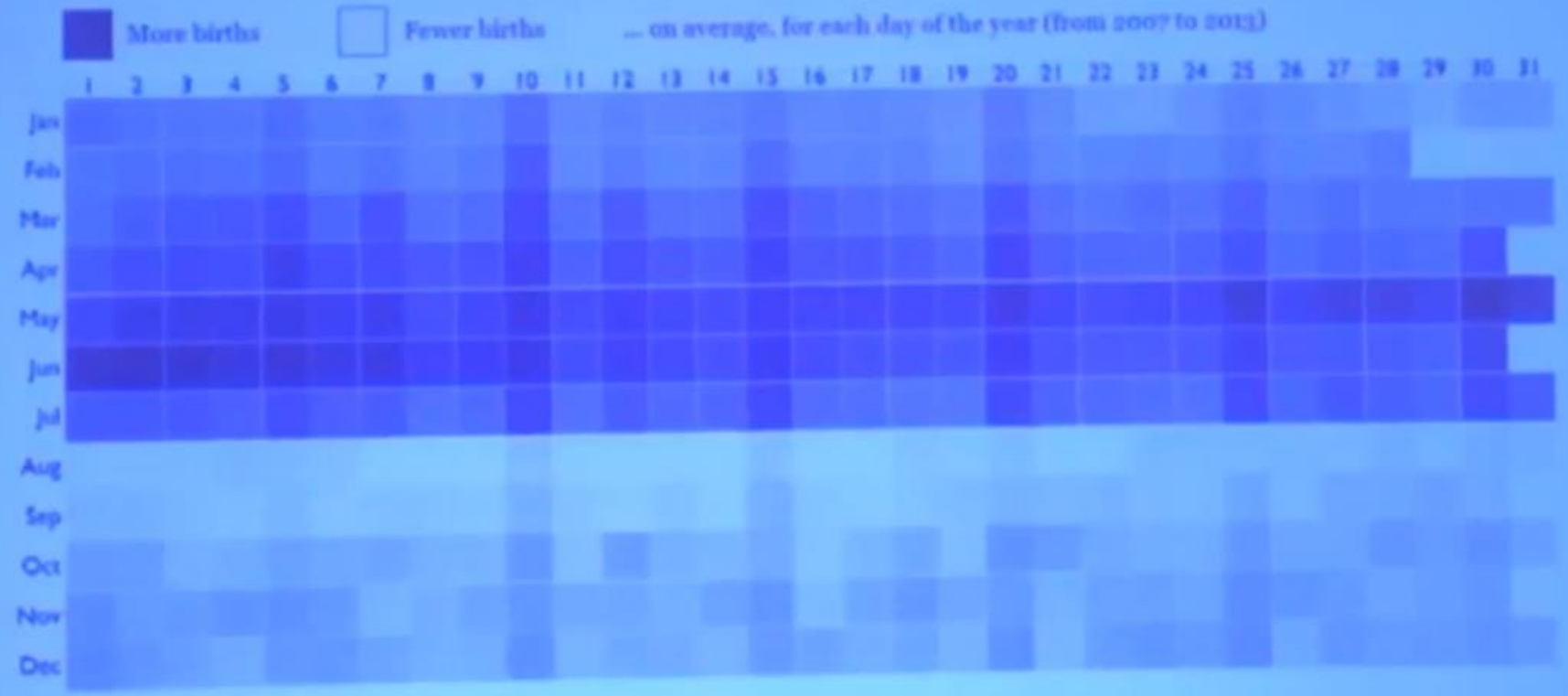
2/29	347th	7/7	10,462
7/5	348th	10/12	10,404
5/26	349th	9/2	10,401
12/31	350th	4/9	10,394
4/13	351st	7/21	10,389
12/23	352nd	4/1	10,338
4/1	353rd	7/9	10,300
11/28	354th	3/7	10,096
11/26	355th	3/5	10,044
11/24	356th	3/3	10,015
10/31	357th	2/7	9,978
11/25	358th	3/4	9,954
11/23	359th	3/2	9,883
11/27	360th	3/6	9,718
12/26	361st	4/4	9,543
1/2	362nd	4/11	9,307
7/4	363rd	10/11	8,796
12/24	364th	4/2	8,069
1/1	365th	4/10	7,792
12/25	366th	4/3	6,574

THE PATTERN IN INDIA IS QUITE DIFFERENT

This is a birth date dataset that's obtained from school admission data for over 10 million children. When we compare this with births in the US, we see none of the same patterns.

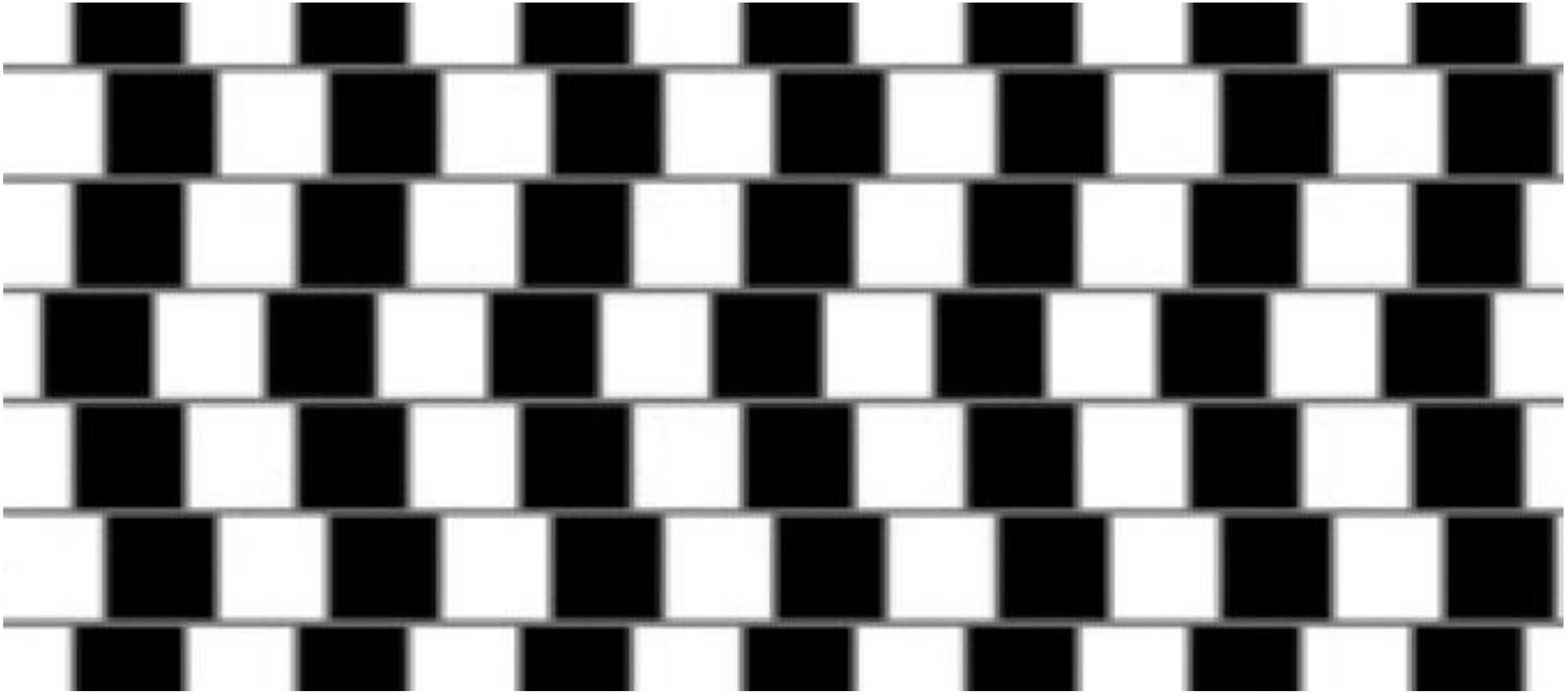
For example,

- Is there an aversion to the 13th or is there a local cultural nuance?
- Are holidays avoided for births?
- Which months have a higher propensity for births, and why?
- Are there any patterns not found in the US data?



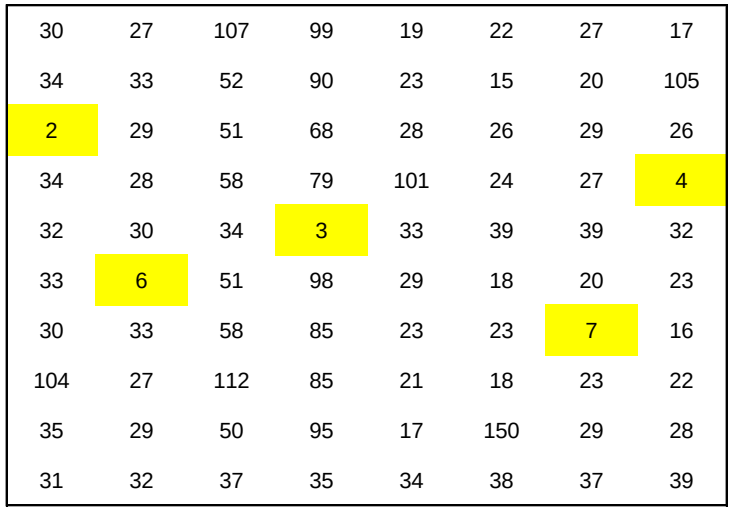
Further Considerations

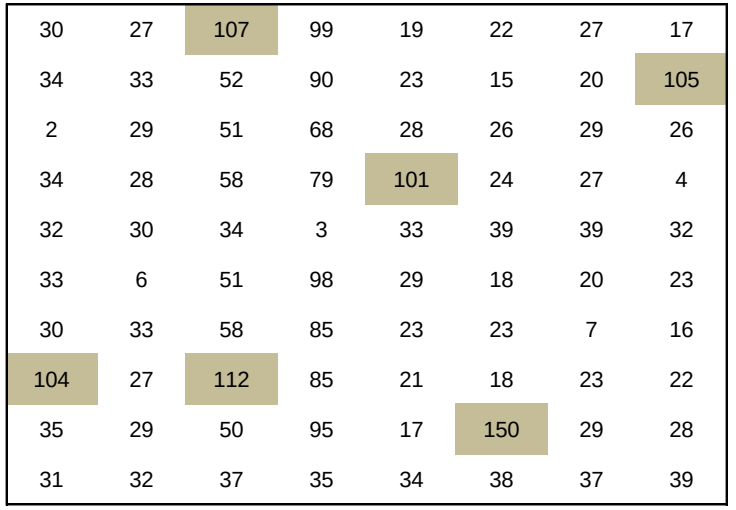
- What data (or subset of data) is relevant for your aim?
- Who is your audience?
- How will you encode your data?
- How will you structure the visualization?
- How will you demonstrate the relationship between data?
- Determine the scale of your visualization.
- Which elements will the user interact with?
- Will a user intuitively understand how the visualization works and what it represents?



How many numbers are less than 10 and Greater than 100?

30	27	107	99	19	22	27	17
34	33	52	90	23	15	20	105
2	29	51	68	28	26	29	26
34	28	58	79	101	24	27	4
32	30	34	3	33	39	39	32
33	6	51	98	29	18	20	23
30	33	58	85	23	23	7	16
104	27	112	85	21	18	23	22
35	29	50	95	17	150	29	28
31	32	37	35	34	38	37	39





Motivation for effective data visualization

- Effective Data Visualization is an art as well as a science
- Focus should be on abstracting out unnecessary data, noise and clutter
- Leverage concepts from the Grammar of Graphics to depict the right information using clean and concise visuals
- "A Picture is worth a thousand words"
- "The greatest value of a picture is when it forces us to notice what we never expected to see." – John Tukey

Tableau

- What is Tableau
- Why should we use Tableau
- How to use Tableau

Introduction to Tableau

- What is Tableau?
  - Tableau is a powerful data visualization software created by Tableau Software.
  - Tableau connects easily to nearly any data source
  - Tableau allows for instant insight by transforming data into interactive visualizations called dashboards.

Why should we use Tableau

	casual development	functionality	mapping	embed in Powerpoint	on-going updates	cost
Xcelsius	very poor	poor (no serious analytics)	limited: provides maps with regions	creates interactive PDF	server connection or download PDF	free for 2008 version; $299/user/1x for server (10 users)
Advizor	good	good	limited: has map background but no lat/long	install local software or call via browser	server connection or download; free reader	$499/user/1x
Spotfire	good	good	can import ESRI shapes, & use in interactive maps; but no lat/long	install local software or call via browser	server connection; paid reader	$948/user/yr or $4,788/user/yr (depends on edition)
QlikView	poor	good	limited: can connect to Google Map	install local software or call via browser	server connection or download; paid reader	not published; $30,000 or more
Lyzasoft	limited	limited	no specific features	install local software or call via browser	server connection; paid workgroup reader	$179/user/year
Tableau	very good	very good	good: auto-adds lat/long to recognized entities	install local software or call via browser	server connection via paid reader; download via free reader	$999/user/1x (stand-alone); $1,000/user for browser access

Different Tableau Versions

1. Tableau Desktop:
It is a self service business analytics and data visualization that anyone can use. It translates pictures of data into optimized queries. With tableau desktop, you can directly connect to data from your data warehouse for live upto date data analysis. You can also perform queries without writing a single line of code. Import all your data into Tableau's data engine from multiple sources & integrate altogether by combining multiple views in a interactive dashboard.

2. Tableau Server:
It is more of a enterprise level Tableau software. You can publish dashboards with Tableau Desktop and share them throughout the organization with web-based Tableau server. It leverages fast databases through live connections.

3. Tableau Online:
This is a hosted version of Tableau server which helps makes business intelligence faster and easier than before. You can publish Tableau dashboards with Tableau Desktop and share them with colleagues.

4. Tableau Reader:
It's a free desktop application that enables you to open and view visualizations that are built in Tableau Desktop. You can filter, drill down data but you cannot edit or perform any kind of interactions.

5. Tableau Public:
This is a free Tableau software which you can use to make visualizations with but you need to save your workbook or worksheets in the Tableau Server which can be viewed by anyone.

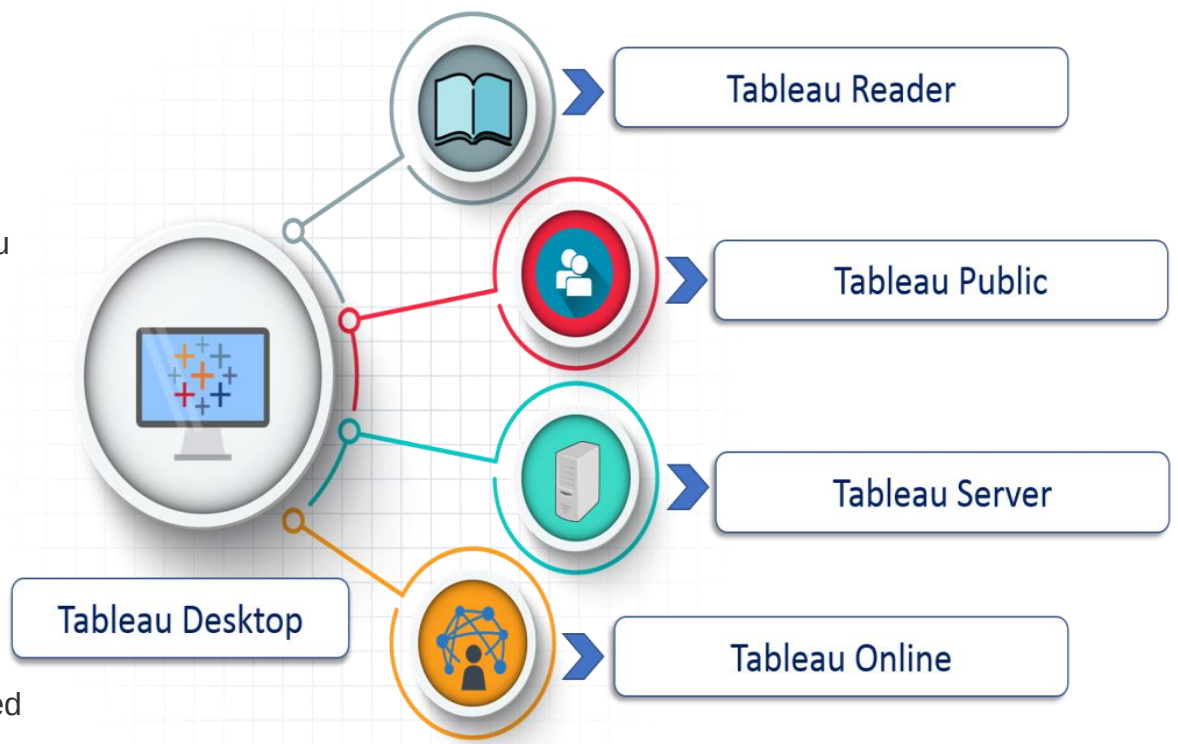
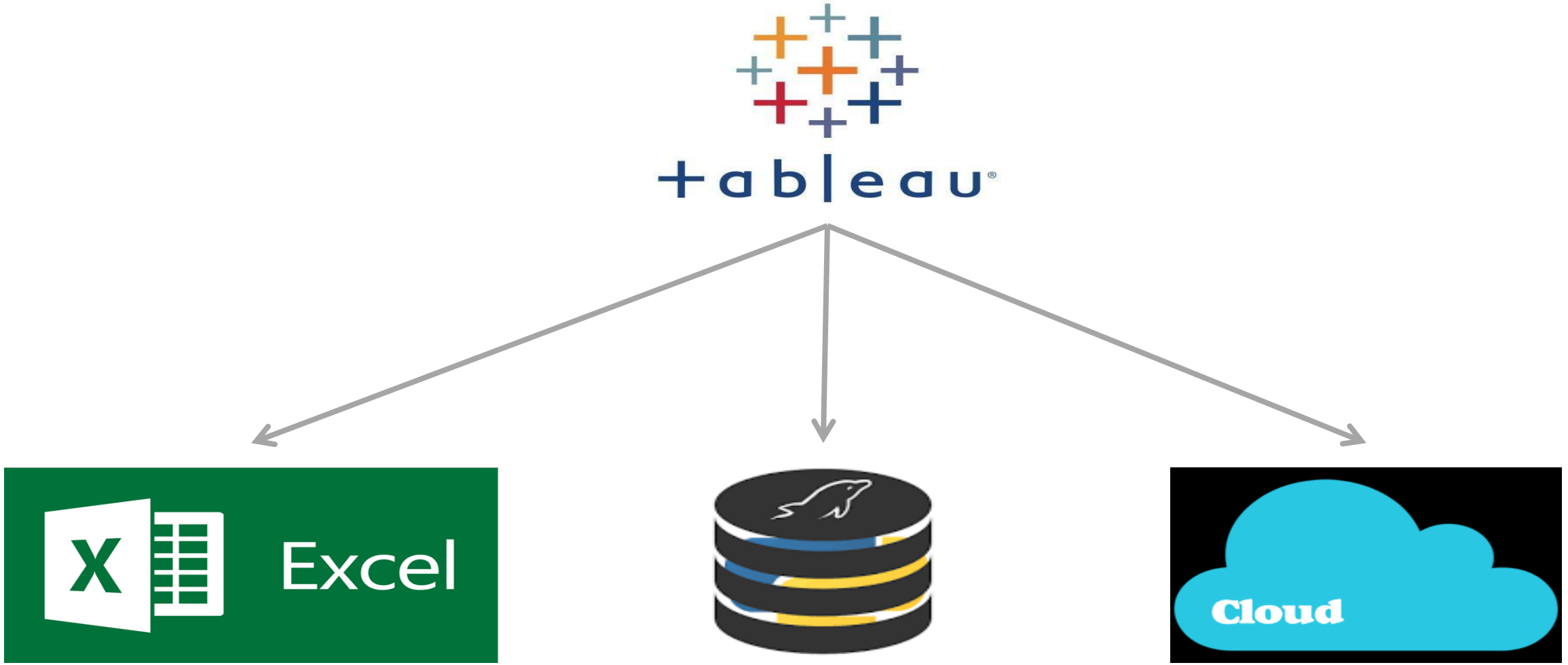
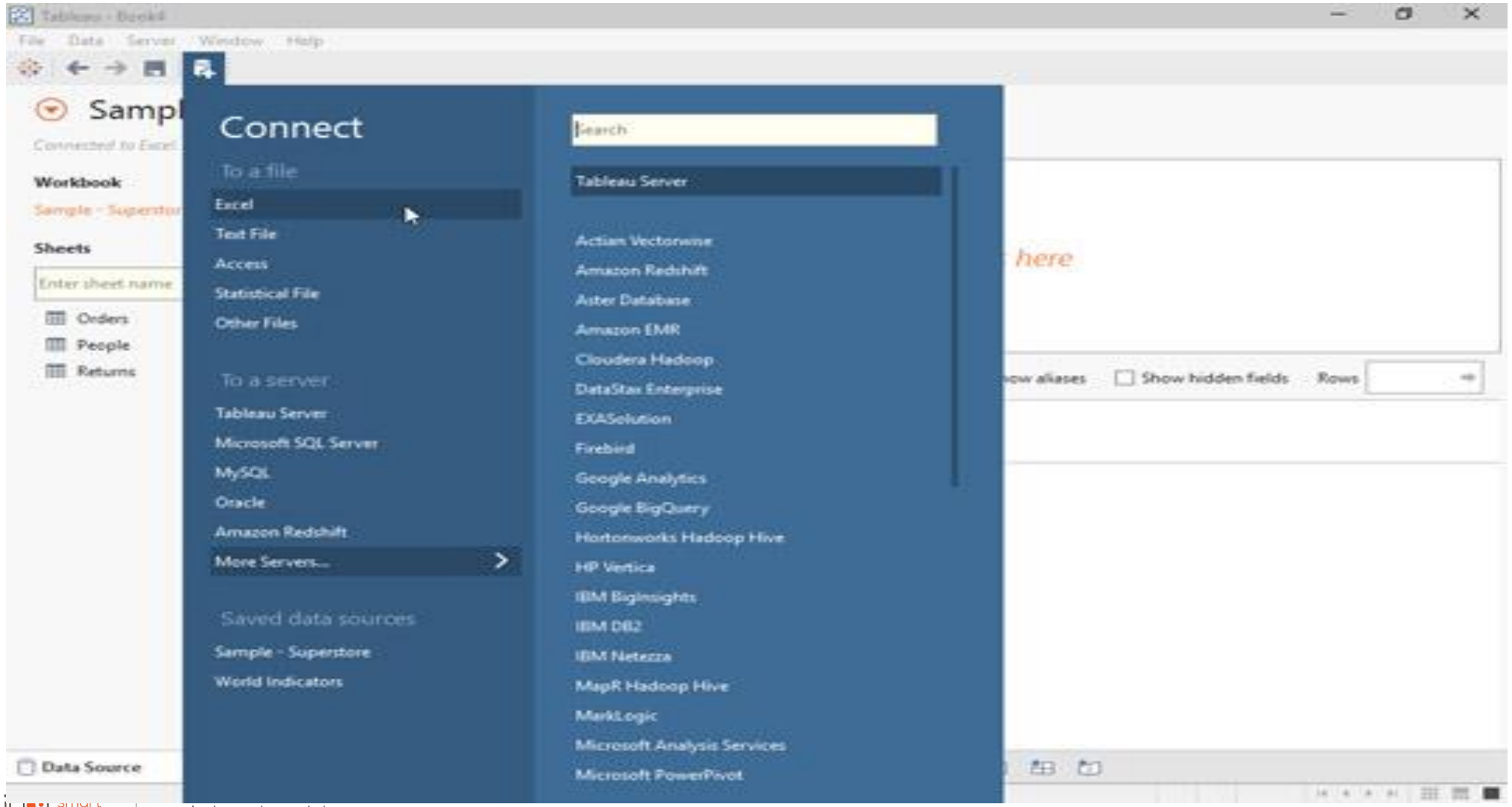


Tableau Features

Connecting to Data

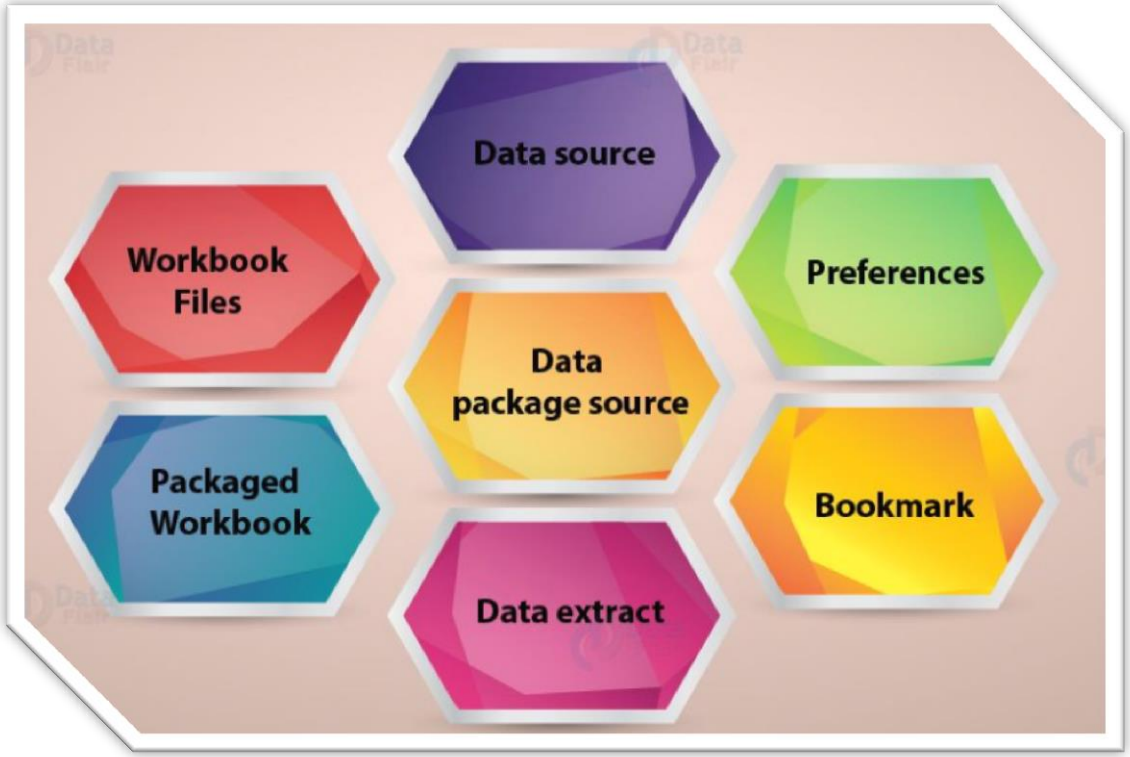


Connecting to Data



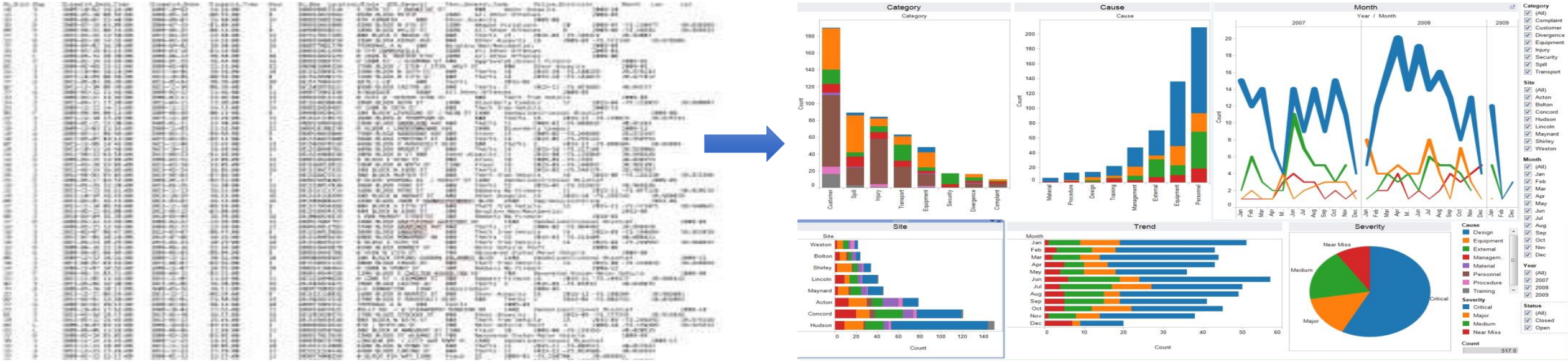
A Note on File types in Tableau

- **Data source(.tds)**: Shortcut to oftenly used data sources, containing information and modifications.
- **Workbook (.twb)**: Workbooks hold worksheets, dashboards and stories.
- **Packaged workbook (.twbx)**: Archive containing a workbook along with all data sources and files.
- **Bookmark (.tbm)**: Bookmarks contain a single sheet.
- **Data extract (.tde)**: Local copy of a subset or entire data source to share and improve the performance.
- **Packaged data source (.tdsx)**: Archive containing a data source file along with any related files.



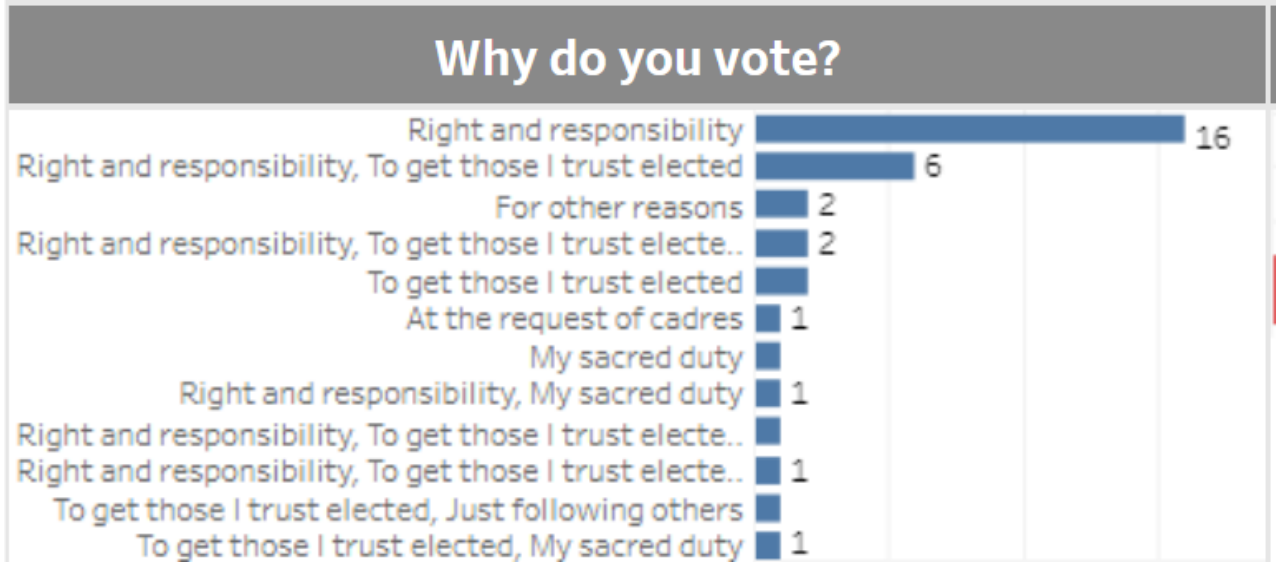
Visual Analytics

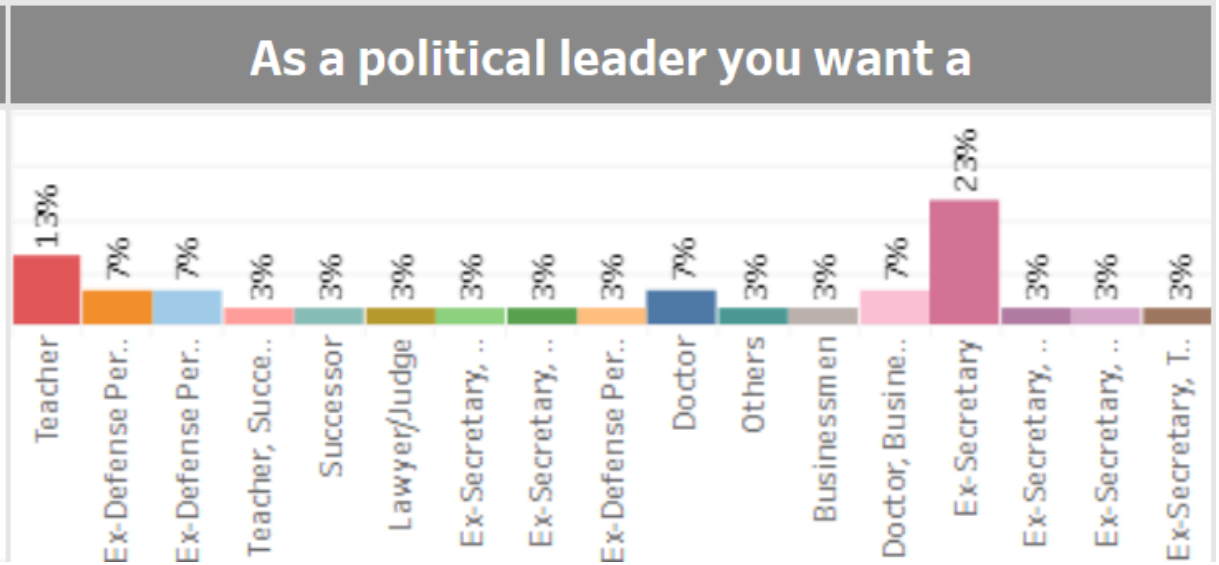
- Is a field in information visualization that focuses on **analytical reasoning** facilitated and supported by interactive **visual interfaces.**



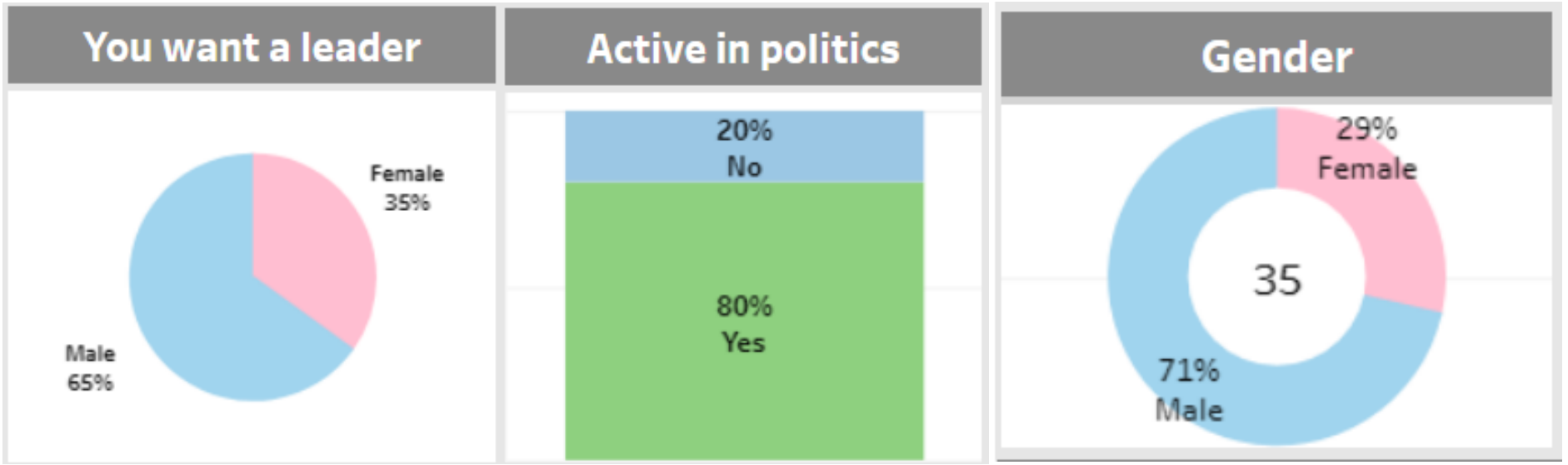
Type of Charts

Simple Bar

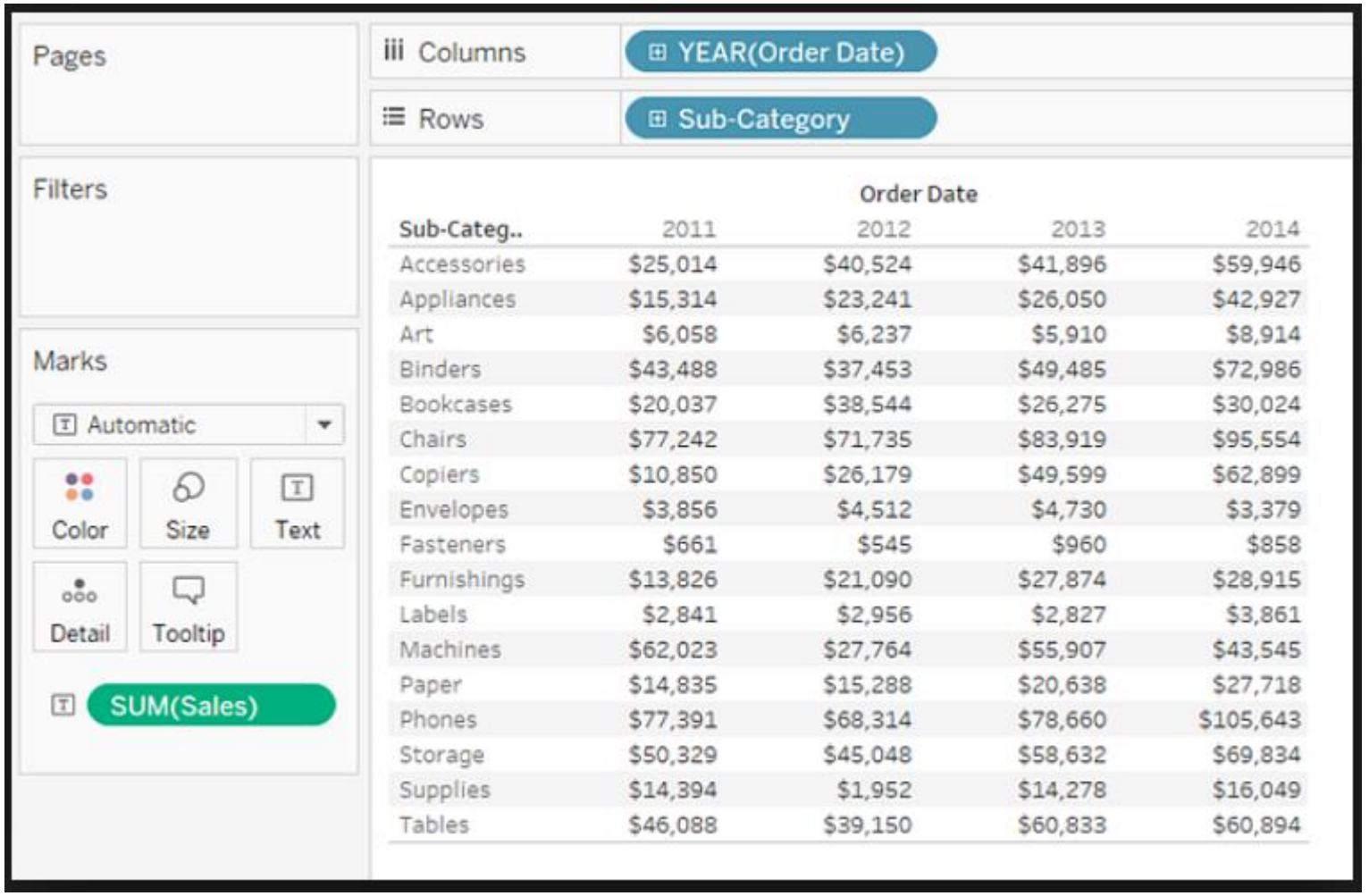

Why do you vote?
Right and responsibility 16
Right and responsibility, To get those I trust elected 6
For other reasons 2
Right and responsibility, To get those I trust electe.. 2
To get those I trust elected
At the request of cadres 1
My sacred duty
Right and responsibility, My sacred duty 1
Right and responsibility, To get those I trust electe..
Right and responsibility, To get those I trust electe.. 1
To get those I trust elected, Just following others
To get those I trust elected, My sacred duty 1


As a political leader you want a
13%
7%
7%
3%
3%
3%
3%
3%
3%
7%
3%
3%
7%
23%
3%
3%
3%
Teacher
Ex-Defense Per..
Ex-Defense Per..
Teacher, Succe..
Successor
Lawyer/Judge
Ex-Secretary, ..
Ex-Secretary, ..
Ex-Defense Per..
Doctor
Others
Businessmen
Doctor, Busine..
Ex-Secretary
Ex-Secretary, ..
Ex-Secretary, ..
Ex-Secretary, T..

Pie, Stacked, Donut

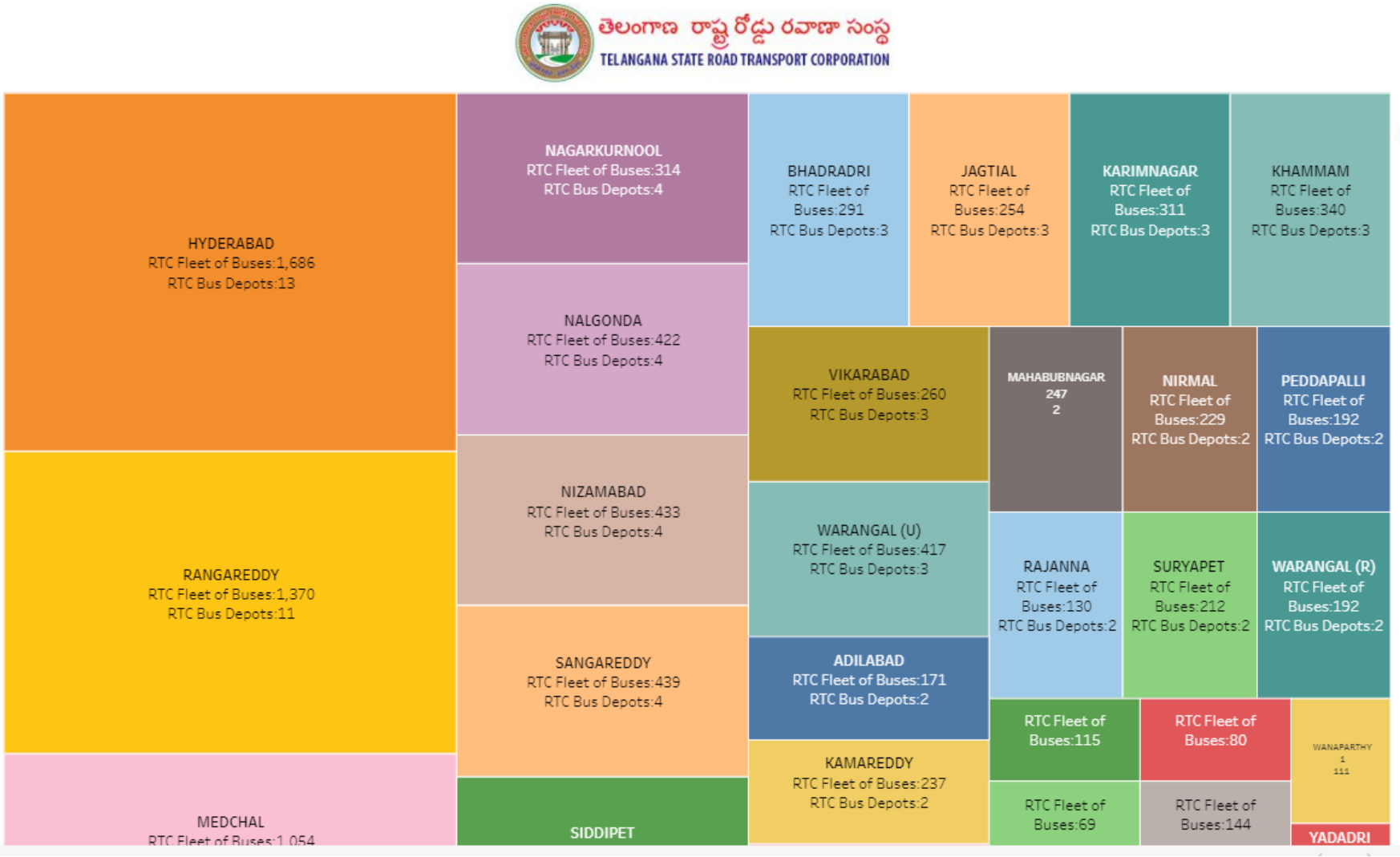


Table



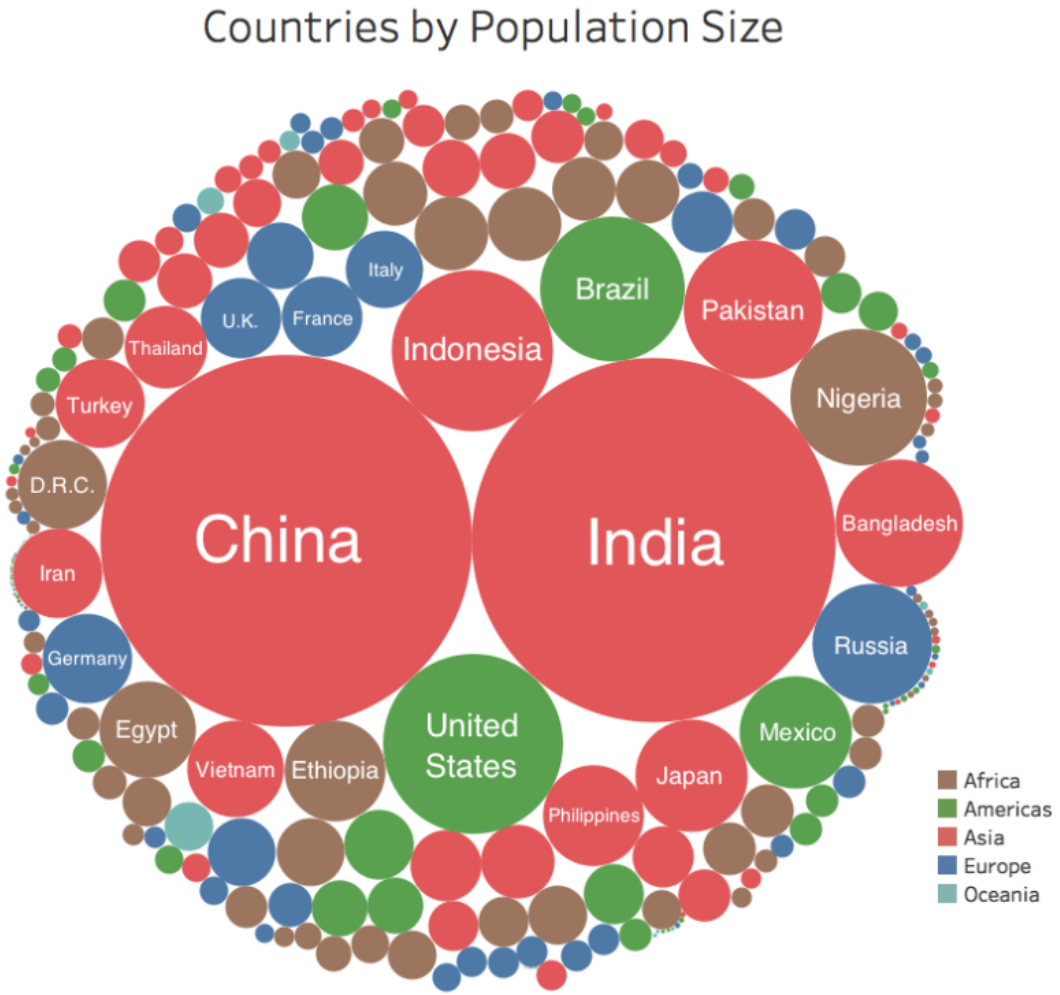
	Order Date			
Sub-Categ..	2011	2012	2013	2014
Accessories	$25,014	$40,524	$41,896	$59,946
Appliances	$15,314	$23,241	$26,050	$42,927
Art	$6,058	$6,237	$5,910	$8,914
Binders	$43,488	$37,453	$49,485	$72,986
Bookcases	$20,037	$38,544	$26,275	$30,024
Chairs	$77,242	$71,735	$83,919	$95,554
Copiers	$10,850	$26,179	$49,599	$62,899
Envelopes	$3,856	$4,512	$4,730	$3,379
Fasteners	$661	$545	$960	$858
Furnishings	$13,826	$21,090	$27,874	$28,915
Labels	$2,841	$2,956	$2,827	$3,861
Machines	$62,023	$27,764	$55,907	$43,545
Paper	$14,835	$15,288	$20,638	$27,718
Phones	$77,391	$68,314	$78,660	$105,643
Storage	$50,329	$45,048	$58,632	$69,834
Supplies	$14,394	$1,952	$14,278	$16,049
Tables	$46,088	$39,150	$60,833	$60,894

Tree map


తెలంగాణ రాష్ట్ర రోడ్డు రవాణా సంస్థ
TELANGANA STATE ROAD TRANSPORT CORPORATION
HYDERABAD
RTC Fleet of Buses:1,686
RTC Bus Depots:13
RANGAREDDY
RTC Fleet of Buses:1,370
RTC Bus Depots:11
MEDCHAL
RTC Fleet of Buses:1,054
NAGARKURNOOL
RTC Fleet of Buses:314
RTC Bus Depots:4
NALGONDA
RTC Fleet of Buses:422
RTC Bus Depots:4
NIZAMABAD
RTC Fleet of Buses:433
RTC Bus Depots:4
SANGAREDDY
RTC Fleet of Buses:439
RTC Bus Depots:4
SIDDIPET
BHADRADRI
RTC Fleet of Buses:291
RTC Bus Depots:3
JAGTIAL
RTC Fleet of Buses:254
RTC Bus Depots:3
KARIMNAGAR
RTC Fleet of Buses:311
RTC Bus Depots:3
KHAMMAM
RTC Fleet of Buses:340
RTC Bus Depots:3
VIKARABAD
RTC Fleet of Buses:260
RTC Bus Depots:3
WARANGAL (U)
RTC Fleet of Buses:417
RTC Bus Depots:3
ADILABAD
RTC Fleet of Buses:171
RTC Bus Depots:2
KAMAREDDY
RTC Fleet of Buses:237
RTC Bus Depots:2
MAHABUBNAGAR
247
2
NIRMAL
RTC Fleet of Buses:229
RTC Bus Depots:2
PEDDAPALLI
RTC Fleet of Buses:192
RTC Bus Depots:2
RAJANNA
RTC Fleet of Buses:130
RTC Bus Depots:2
SURYAPET
RTC Fleet of Buses:212
RTC Bus Depots:2
WARANGAL (R)
RTC Fleet of Buses:192
RTC Bus Depots:2
RTC Fleet of Buses:115
RTC Fleet of Buses:80
WANAPARTHY
1
111
RTC Fleet of Buses:69
RTC Fleet of Buses:144
YADADRI

Bubble

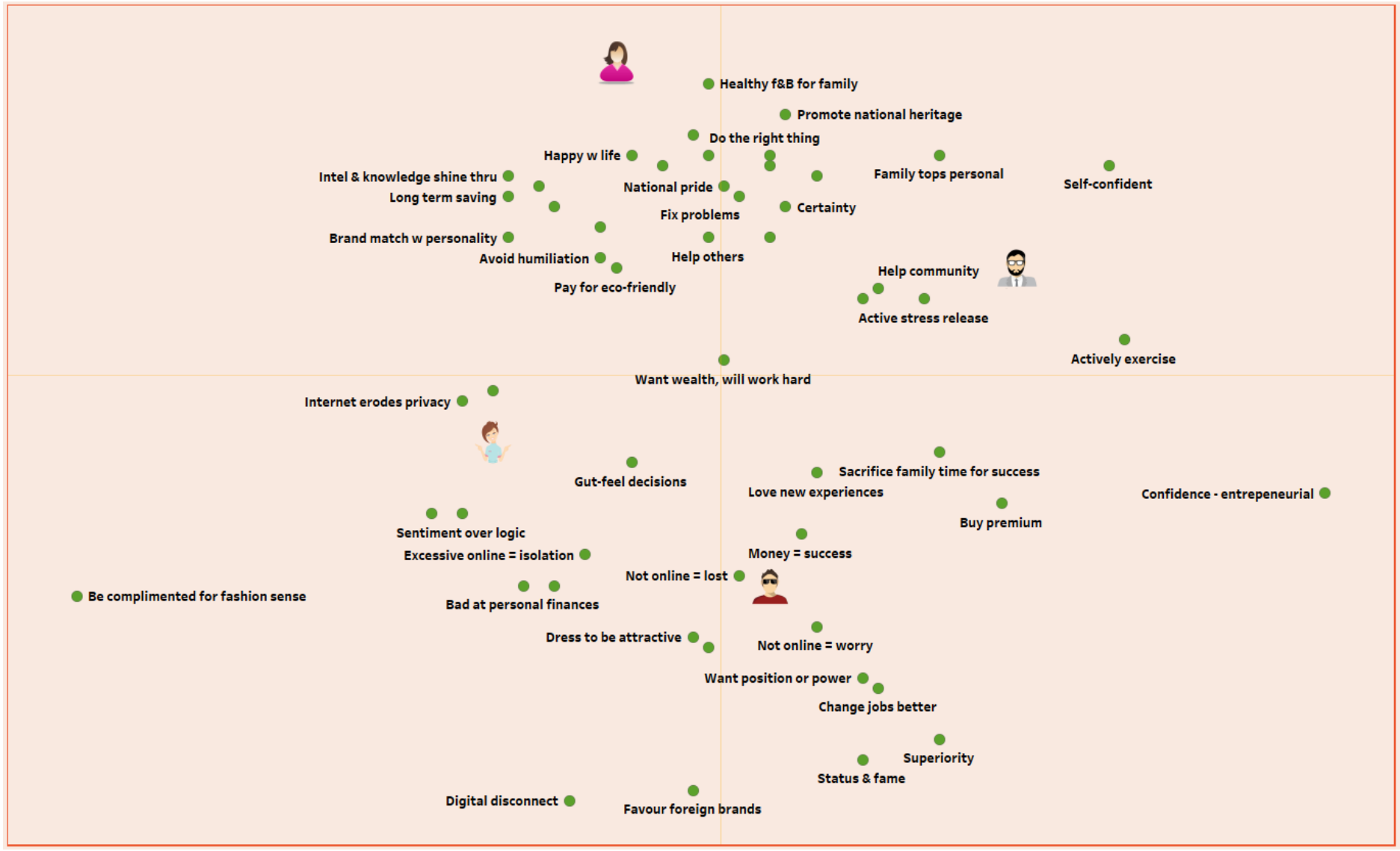
It’s a simple visualization, but the story that gets distilled is loud and clear:



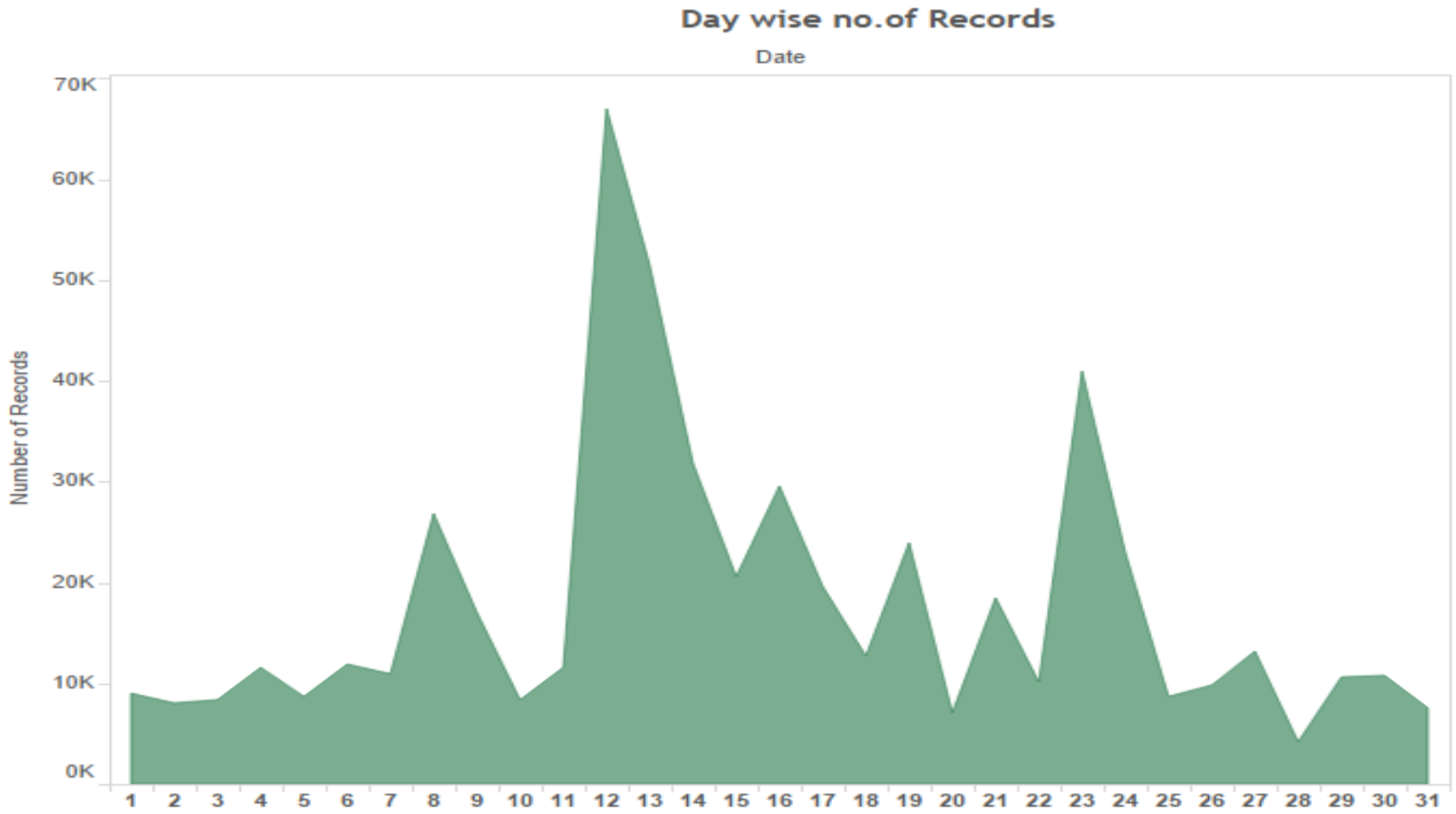
Word Cloud

The image below shows a sample word cloud of 100 most used passwords. One can easily interpret that “123456” is most used password as represented by its size followed by “password” followed by“12345678” and so on.

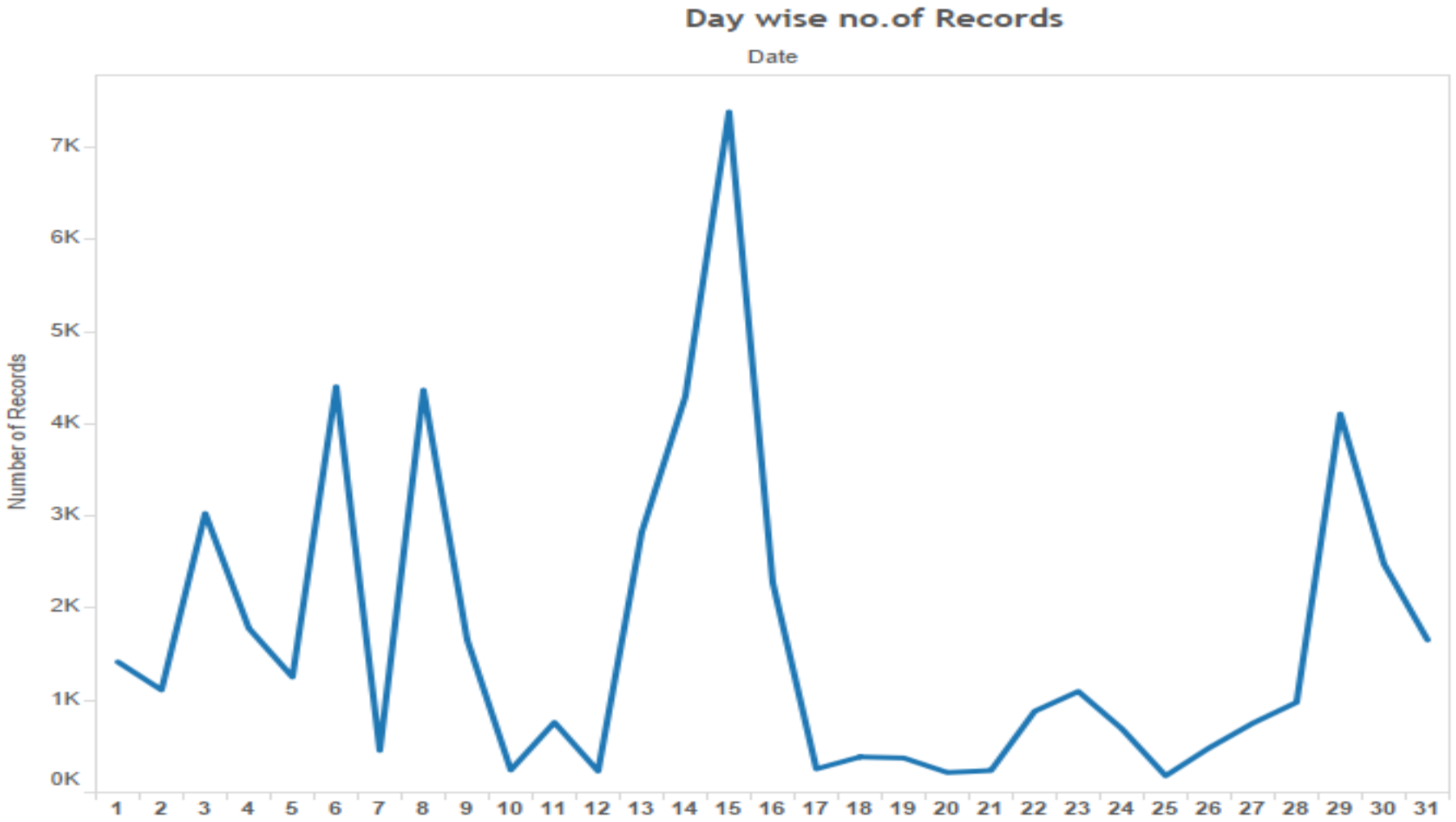
2 x 2


Healthy f&B for family
Promote national heritage
Do the right thing
Happy w life
Intel & knowledge shine thru
Long term saving
National pride
Family tops personal
Self-confident
Certainty
Fix problems
Brand match w personality
Avoid humiliation
Help others
Pay for eco-friendly
Help community
Active stress release
Actively exercise
Want wealth, will work hard
Internet erodes privacy
Gut-feel decisions
Sacrifice family time for success
Love new experiences
Confidence - entrepeneurial
Buy premium
Sentiment over logic
Excessive online = isolation
Money = success
Not online = lost
Be complimented for fashion sense
Bad at personal finances
Dress to be attractive
Not online = worry
Want position or power
Change jobs better
Superiority
Status & fame
Digital disconnect
Favour foreign brands

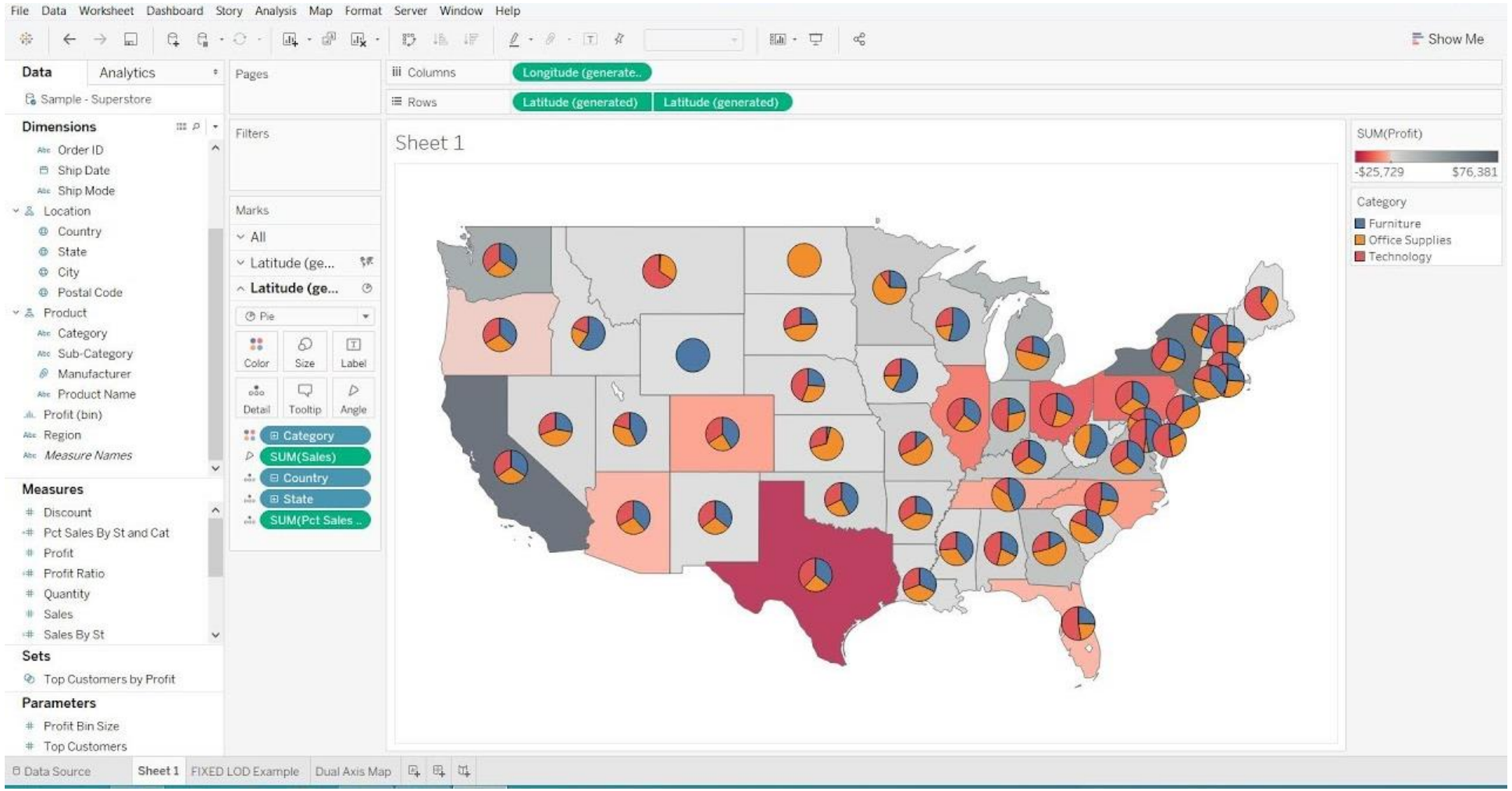
Area


Day wise no.of Records
Date
Number of Records
70K
60K
50K
40K
30K
20K
10K
0K
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Line


Day wise no.of Records
Date
Number of Records
7K
6K
5K
4K
3K
2K
1K
0K
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

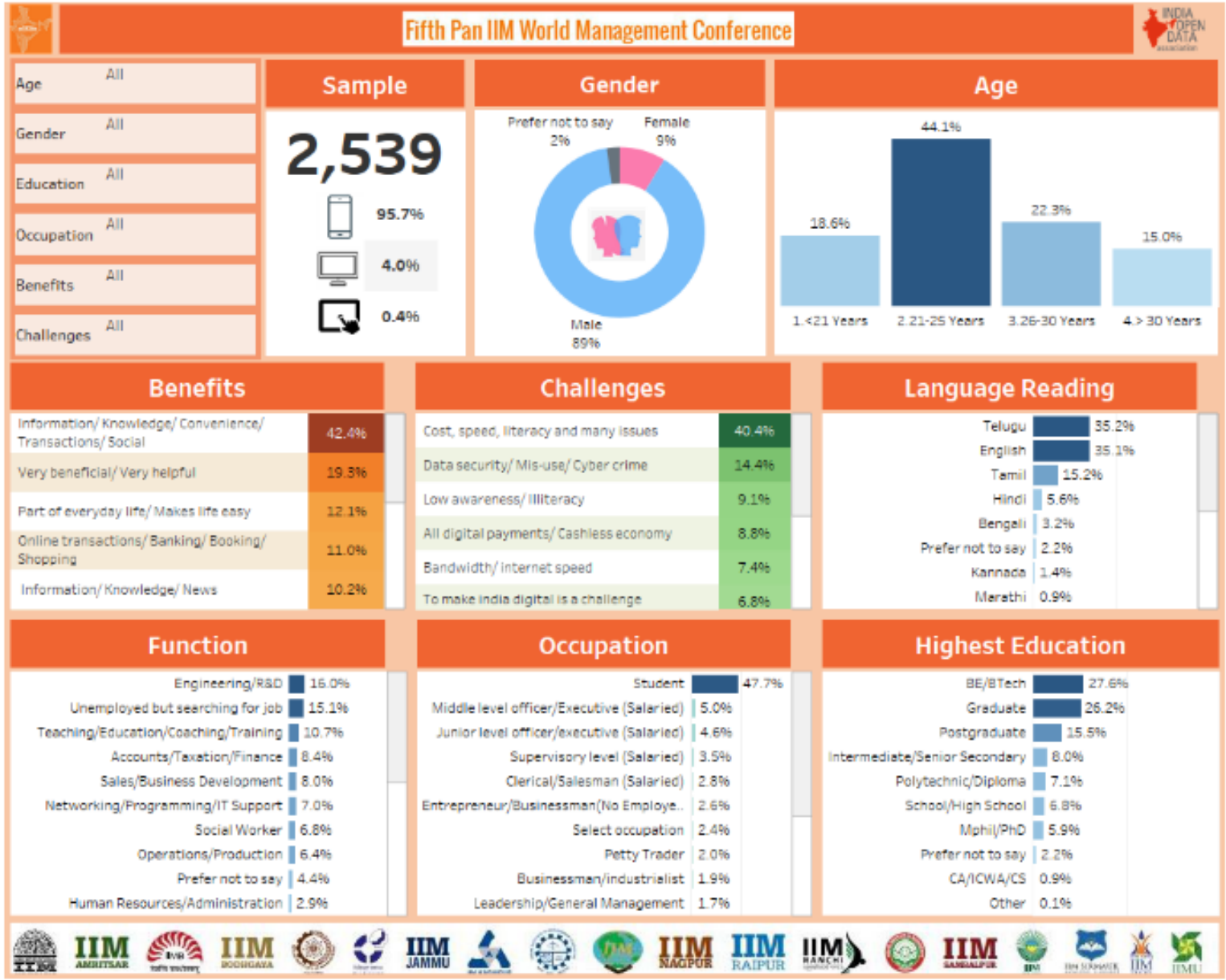
Heat Map


File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Sample - Superstore
Dimensions
Order ID
Ship Date
Ship Mode
Location
Country
State
City
Postal Code
Product
Category
Sub-Category
Manufacturer
Product Name
Profit (bin)
Region
Measure Names
Measures
Discount
Pct Sales By St and Cat
Profit
Profit Ratio
Quantity
Sales
Sales By St
Sets
Top Customers by Profit
Parameters
Profit Bin Size
Top Customers
Pages
Filters
Marks
All
Latitude (ge...
Latitude (ge...
Pie
Color
Size
Label
Detail
Tooltip
Angle
Category
SUM(Sales)
Country
State
SUM(Pct Sales ..
Columns
Longitude (generate..
Rows
Latitude (generated)
Latitude (generated)
Sheet 1
SUM(Profit)
-$25,729
$76,381
Category
Furniture
Office Supplies
Technology
Data Source
Sheet 1
FIXED LOD Example
Dual Axis Map

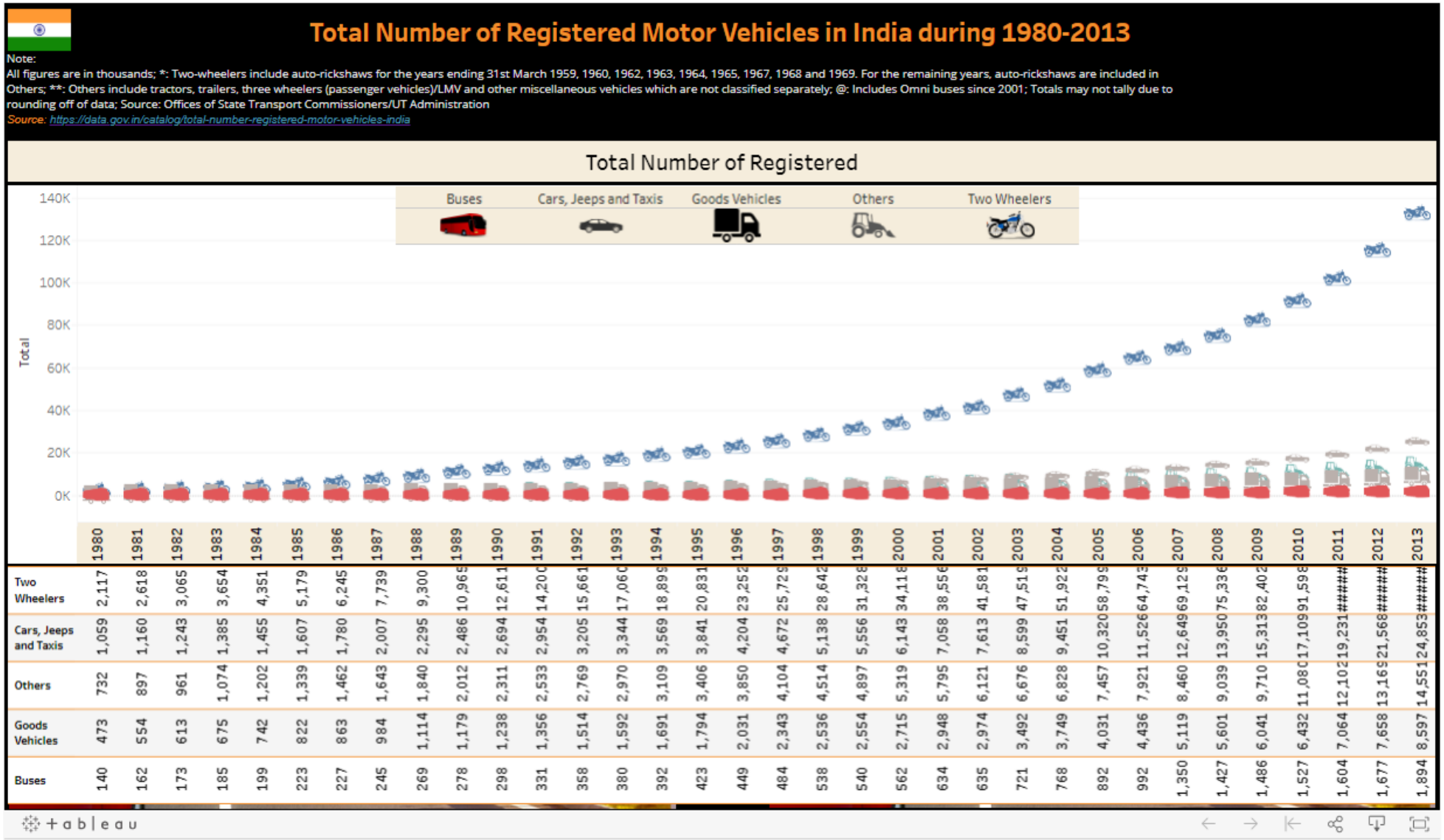
Boxplot


Pages
Columns
Segment
Region
Rows
Discount
Filters
Marks
Circle
Color
Size
Label
Detail
Tooltip
Segment / Region
Consumer
Corporate
Home Office
Discount
80%
70%
60%
50%
40%
30%
20%
10%
0%
Central
East
South
West

Click on picture to view on web



Click on picture to view on web



Click on picture to view on web

Reset Filter

Click on the picture to leave all filters if we selected any in the light grey box.

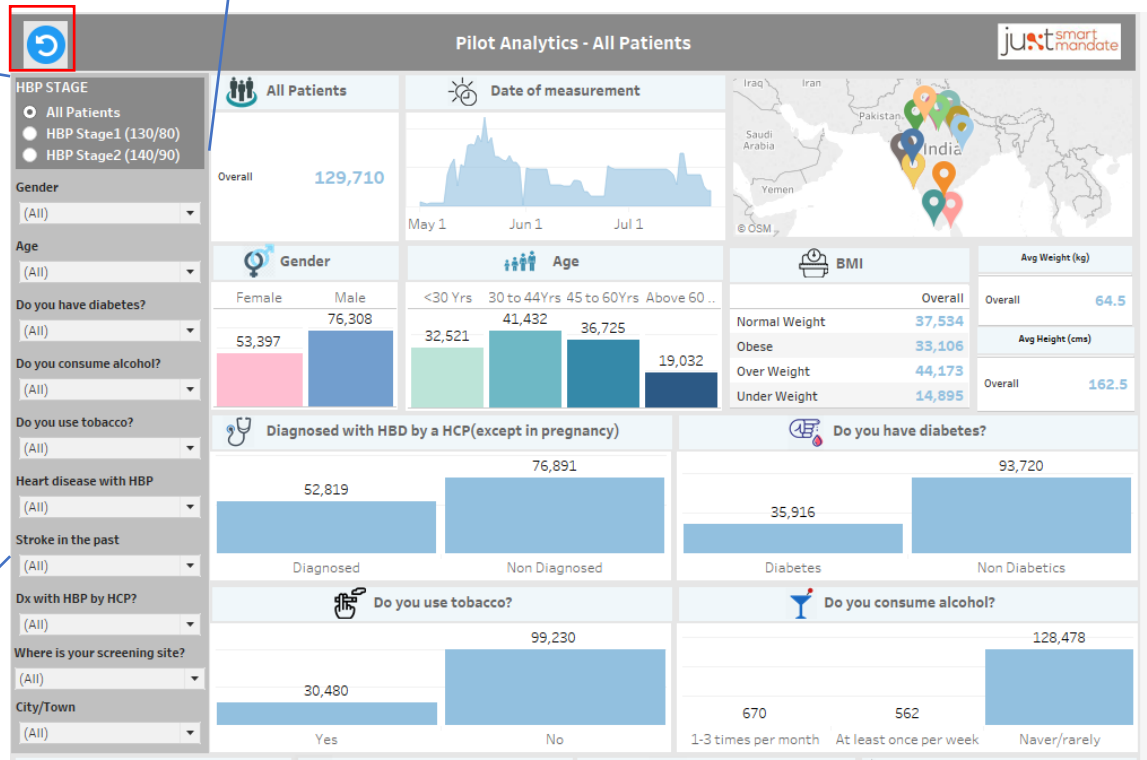
Dark Grey

A **parameter** will allow you to provide a value to pass into **Tableau**. **Parameters** allo w you to come up with scenarios or options that are not available in your data and create these values to put into your visualization.

We can see our dashboard by 3 type of patients i.e.
1.Overall
2.HBP Stage1
3.HBP Stage2

Light Grey

Filter restricts the number of records present in data set based on given condition/selection



Simple sample Dashboard

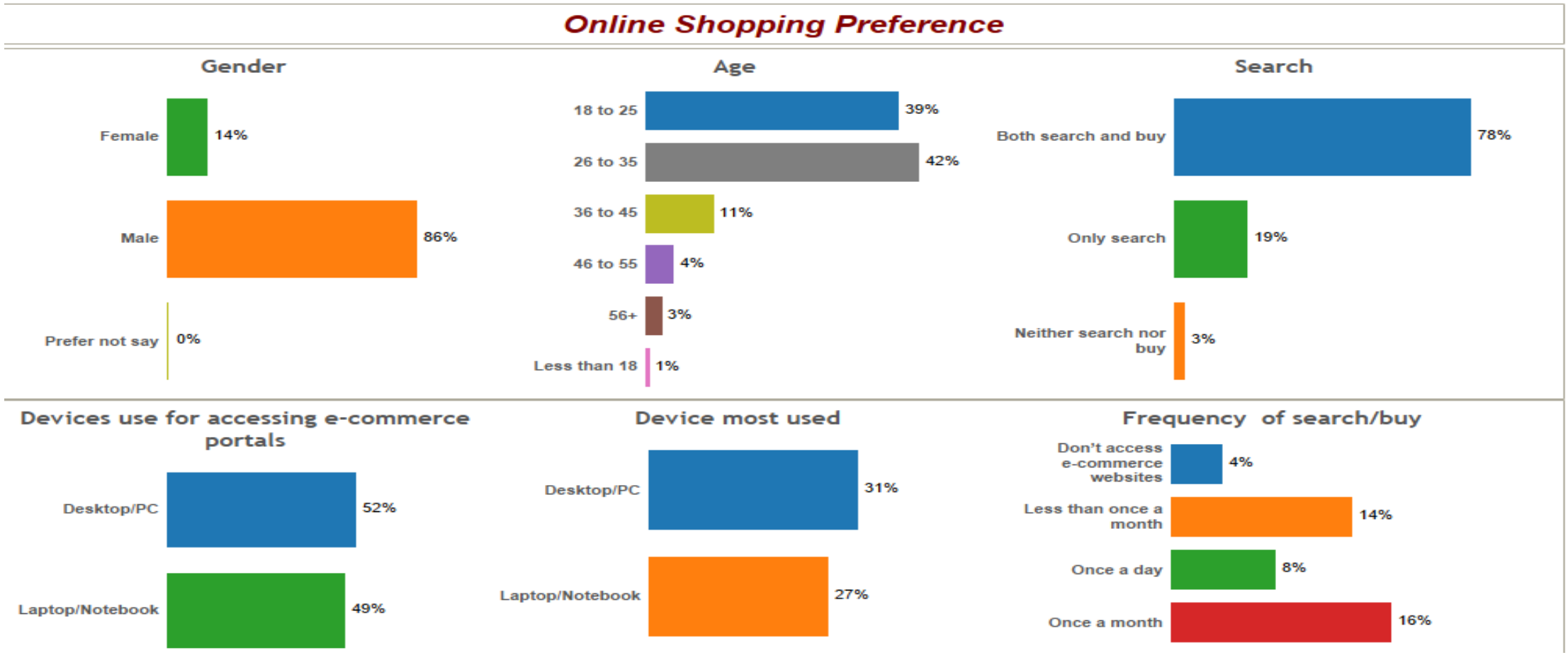


Chart Suggestions—A Thought-Starter