What is Tableau and Why Should I Care?

Karen Rahmeier and Melissa Perry, Codecinella

Madison WI, June 26, 2018

About me – Karen Rahmeier

- Software developer since 1998
- Team Lead of software developers, Wisconsin Technical College System
- Design and build full-stack, multi-tier applications using the Java Enterprise stack for various State of Wisconsin agencies
- One of the organizers of Codecinella

About me – Melissa Perry

- Six years work experience in automotive, education and telecommunications industries, all involving using data for action
- Transitioning into data science role from past roles in info management and visualization
- Experience in Helping decision makers understand the truth (as told by data) and act effectively

Karen’s technology stack

- Currently using:
  - Microsoft SQL Server (DB, SSRS, SSIS, SSAS, Visual Studio)
  - Java (Eclipse, Struts, Glassfish, jBoss, IntelliJ)
  - (SQL, HTML, javascript, css, .jsp, xml)
  - Tableau (server and desktop)
  - Subversion, Jira (Git, Confluence)

Melissa’s technologies and skills

- Currently using:
  - Database tools (SQL Server, Postgres, Oracle)
  - Analytics tools (R, SAS, SPSS, STATA)
  - Red Hat Linux
  - Tableau (server and desktop)
  - Process tools (Jira, Git)
  - Root cause problem solving
  - Results based accountability

Melissa's motivation:
To connect non-techies with the Truth

- Overly technical speak intimidates others; let's stop this.
- Tableau helps bridge the gap.
- We will always be accountable for effective action.
- Your boss/customer doesn't [need to] understand everything you're doing to make the decision; they do need to understand what you think the decision should be.

Karen's motivation:

Could Tableau do what machine learning or predictive analytics do?

- Somewhat intimidated by very technical talks
- Visual learner
- Creative thinker and learner
  - How did Naveen's words fit in the world and skills I know?
  - How could I use her concepts in my real job?
  - Should I learn/ could I use what she was talking about?
  - What technical skills should I next develop?

Tableau at Karen's job

- New person hired
- Tableau introduced 2-3 years ago
- Slowly identifying how to use it at work
- Move away from static reports, 'cubes'
- Increase emphasis on 'data' as a 'value-add'
- First Tableau-centric project launched

- … it's time to learn Tableau

Goals for today

- Awareness of what Tableau software is and how it should and should NOT be used
- The value of Tableau and data visualization
- See Tableau demonstrated in action and get a sense of how you could use it in real life

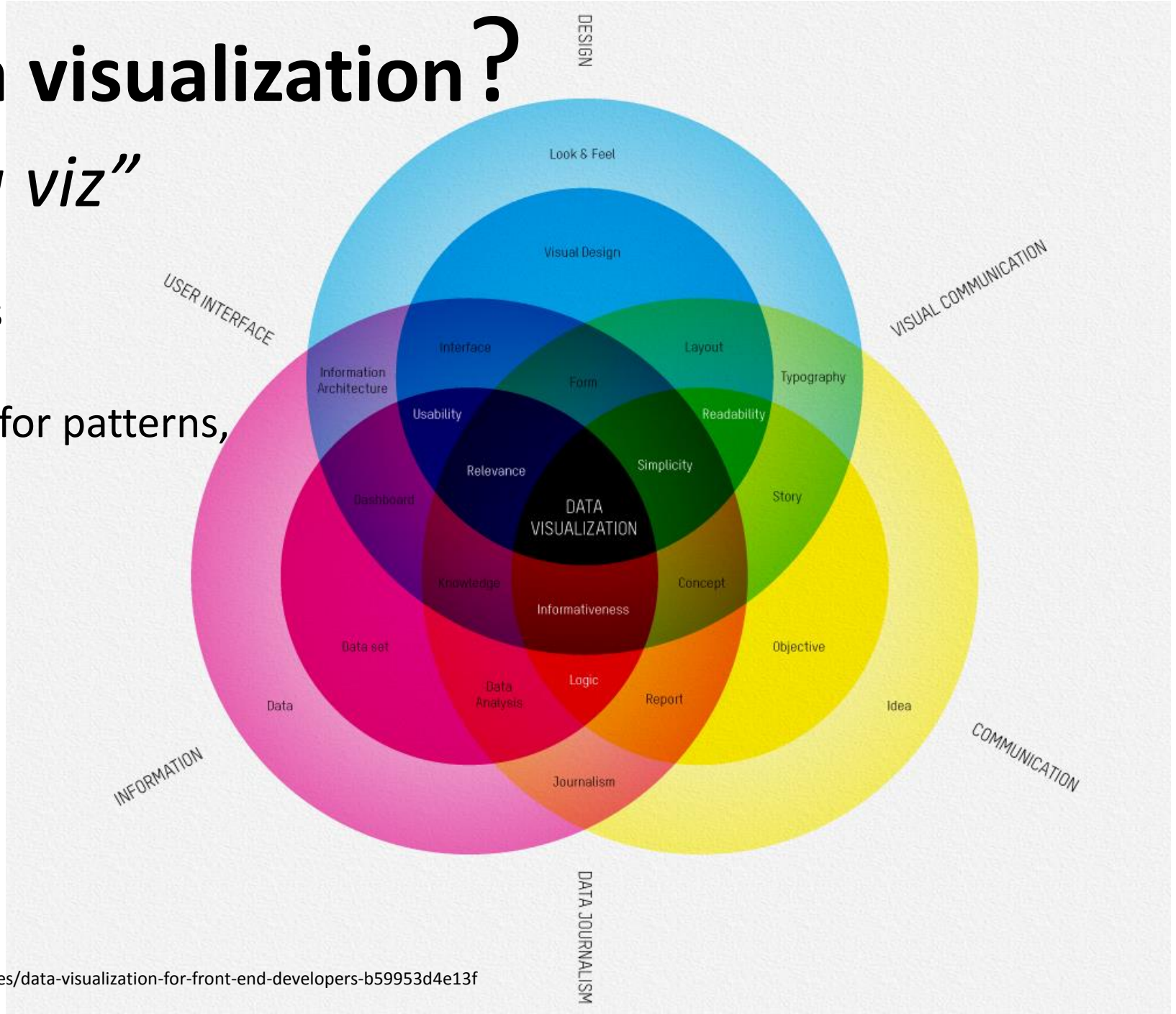
Outline (how we'll accomplish goals)

- Tableau in context
  - What is data visualization and how does it touch common technology careers in business intelligence, data science, and software engineering?
  - Industry domains most actively using data viz
  - Tableau compared to other data viz tools
- Motivation for using data viz in general
- Why Tableau is our first choice as a tool
- Relevancy to previous Codecinella talks
- Demo

What is **data visualization**?

a.k.a *"data viz"*

- Dashboards
- Infographics
- Stories
- Exploration for patterns, trends



Source: https://medium.com/@ksykes/data-visualization-for-front-end-developers-b59953d4e13f

Source: http://tclive.tableau.com/Library/Video?vCode=17BC-001

Source: http://tclive.tableau.com/Library/Video?vCode=17BC-001

Professionals in all these careers use data viz through different work output:

- Business Intelligence
  - Answering questions about data that is already understood (performance/accountability)
  - Ex. Might design a common dashboard for the Sales team to understand performance
- Data Science
  - Answering deeper questions, bringing new insight, develop algorithms that solve problems with ‘big data’
  - Ex. Might use a data viz story about selection of the appropriate data model
- Software Engineering
  - Designing, developing, and delivering a piece of software as specified by a customer (think app’s that collect data, databases and infrastructure, computing infrastructure)
  - Ex. Might embed a chart inside an app to engage customers and encourage use, like Fitbit.com

How People Use Tableau

Industries | Departments | Capabilities | Technology

Communications, Media, & Technology
Communications & Media Analytics

High Technology Analytics

Sports Management Analytics

Energy & Resources
Energy & Resources Analytics

Financial Services
Banking Securities & Investments Analytics

Insurance Analytics

Healthcare & Life Sciences
Healthcare Analytics

Life Sciences Analytics

Manufacturing
Manufacturing Analytics

Public Sector
Associations & Non-profits Analytics

Education Analytics

Government Analytics

Retail & Consumer Goods
Consumer Goods Analytics

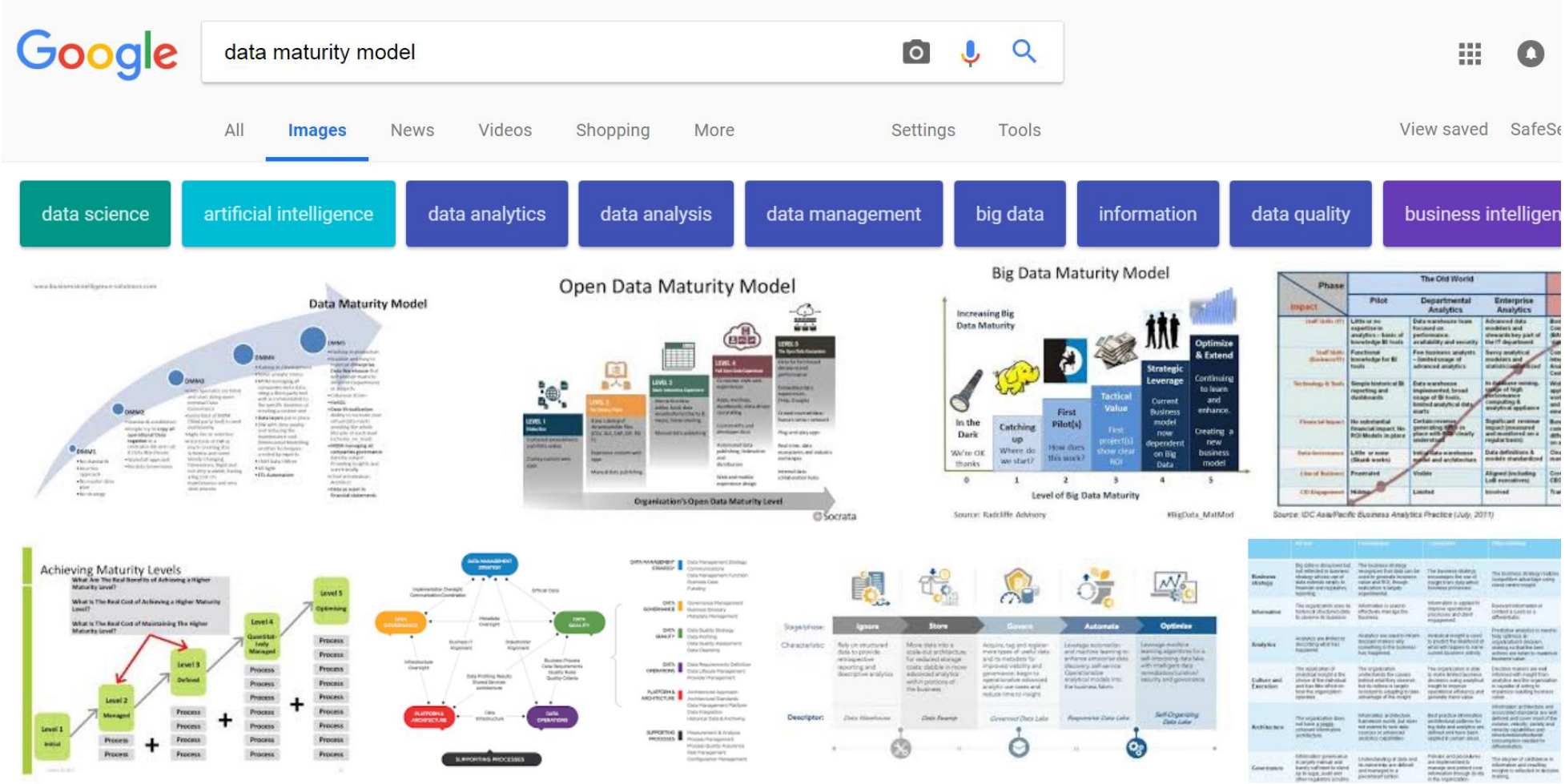
Retail & Wholesale Analytics

Services
Services Analytics

Travel & Transportation
Travel & Transportation Analytics

Source: https://www.tableau.com/solutions

Assess the data maturity of each industry/sector.



Source: https://optimalbi.com/blog/2017/02/17/gartner-magic-quadrant-for-business-intelligence-2017-cloud-is-coming-slowly/

Adaptive Challenges with Tableau

(Adaptive Challenges, unlike technical ones,
have no known solutions that will work in every case.)

- Technical enough to ‘count’, accessible enough to learn
- Karen’s job: the most brilliant and technical person introduced Tableau, therefore rapid acceptance and adoption
- Melissa’s job: Tableau available alongside competitor’s tool, seen as non-technical/easy, and sometimes toil, culture with low exposure to UX philosophy

Remember

1) **Aesthetics matter.**
2) **Results matter.**

Adaptive Challenges with Tableau

(Adaptive Challenges, unlike technical ones,
have no known solutions that will work in every case.)

“Today’s businesses do not rely upon an analytical tool just because it’s latest in the market.

They want advanced, convenience (sic) and smart tools that help analyze complex data instantly.

Gone are the days when the single -dimensional and static graphs and charts were used to demonstrate the statistics.

Businesses need more visually appealing, interactive and agile tools that fit into the real-time requirements.”

https://intellipaat.com/blog/tableau-vs-qlikview-difference/

Past Codecinella tech talks…

..that got Karen thinking about data visualization

- Machine Learning (January 2018)
- Predictive Analytics (January 2017)

Machine Learning

MACHINE LEARNING

Posted on January 23rd, 2018

Naveen VK presented on machine learning and how artificial intelligence works. She

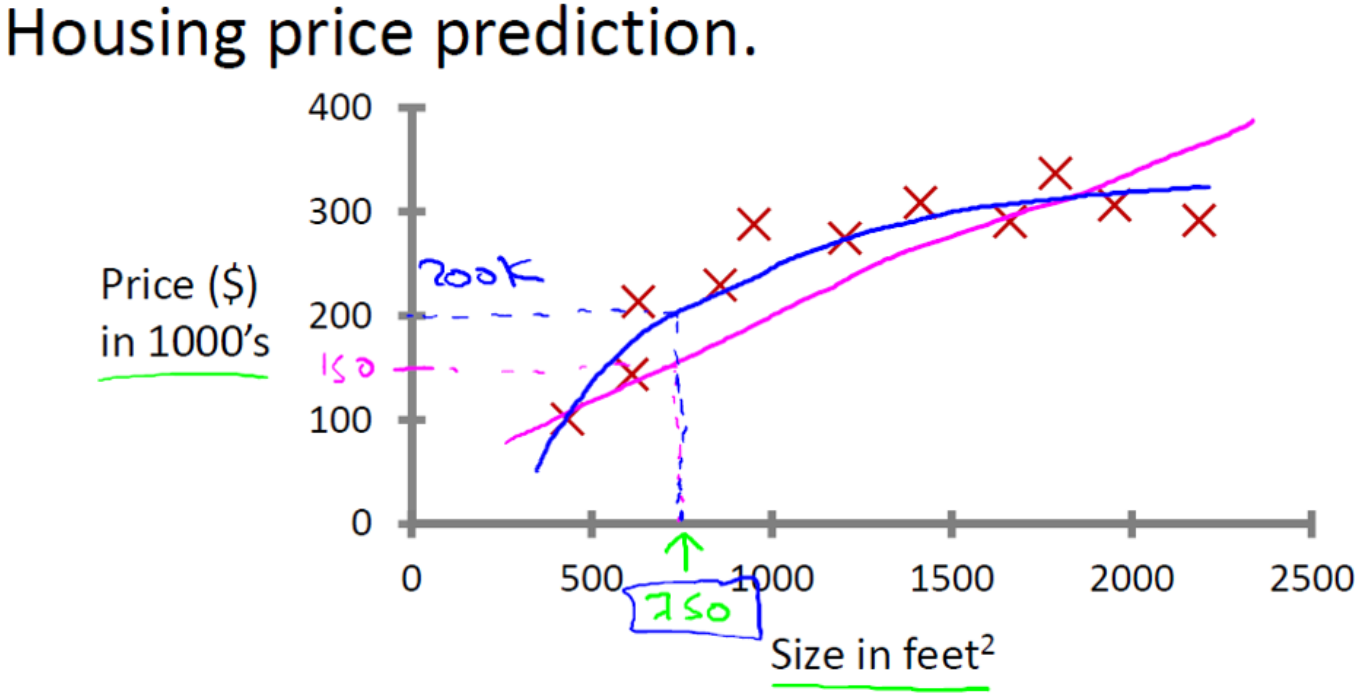
Disclaimer and Acknowledgement

- **Presentation based on Machine Learning Algorithms class on coursera.org by Stanford University professor Dr. Andrew Ng**

https://www.coursera.org/learn/machine-learning

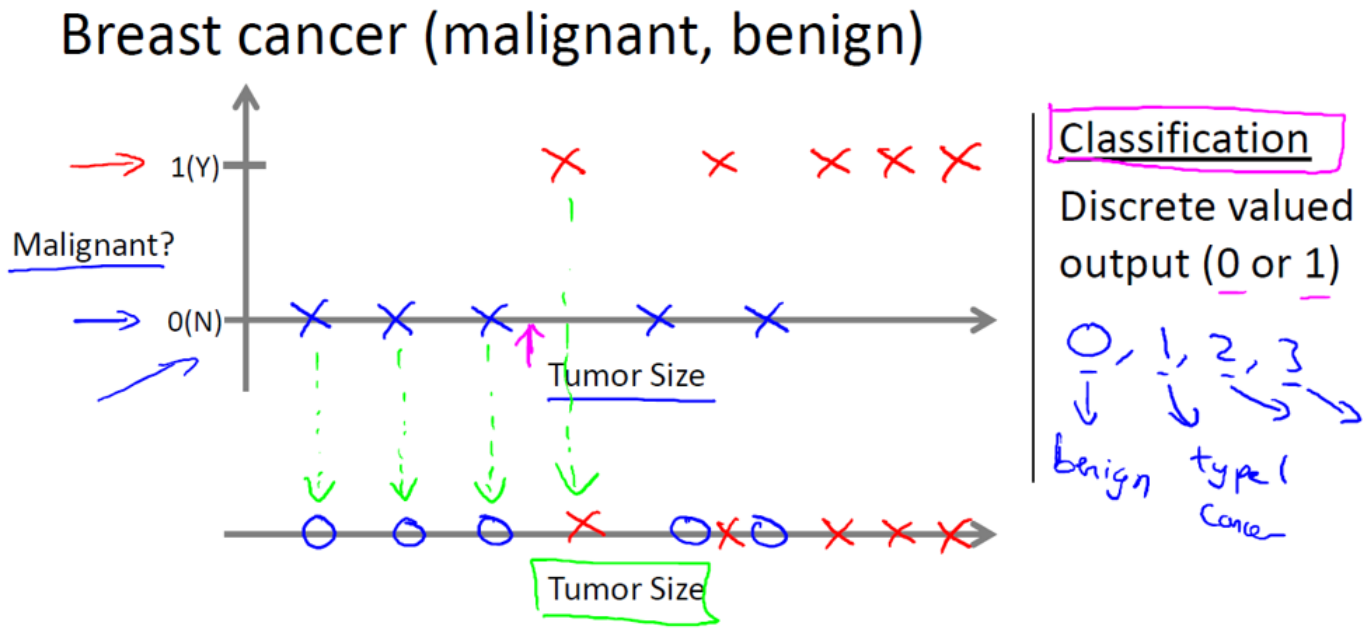
Machine Learning … visual data (1)

Supervised Machine Learning: Regression
Example: House Price Prediction



Machine Learning … visual data (2)

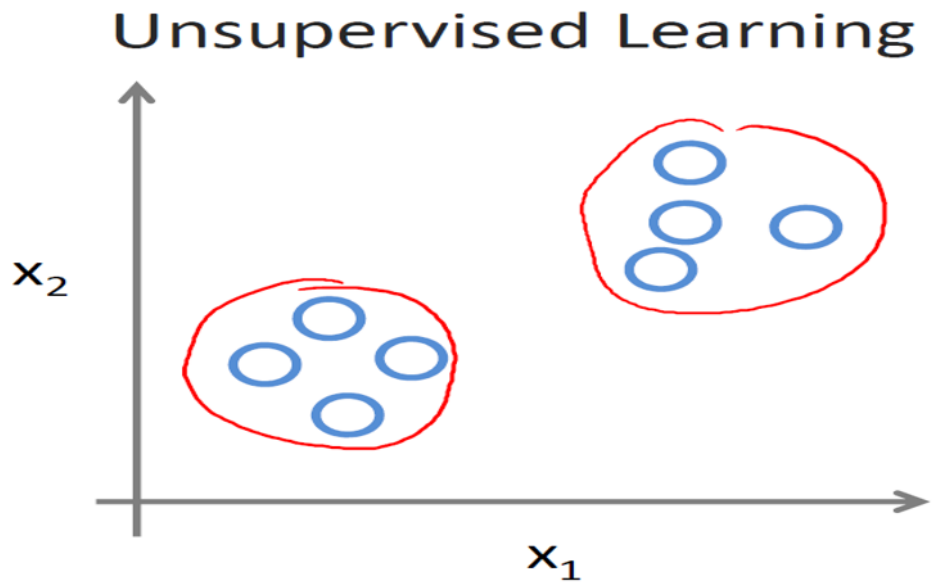
Supervised Machine Learning: Classification
Example: Breast cancer prediction



Machine Learning … visual data (3)

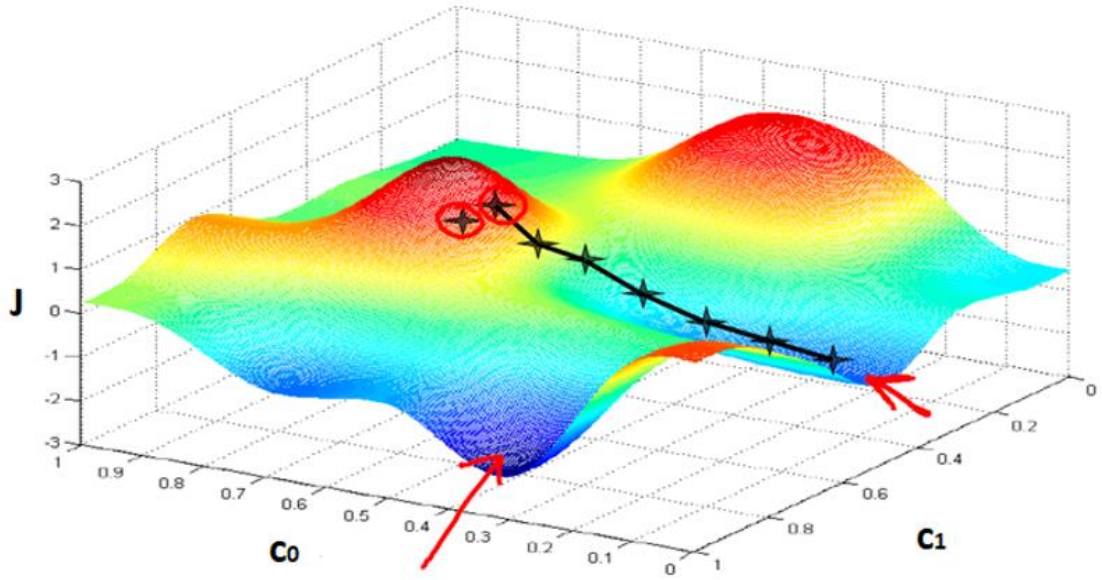
Unsupervised Machine Learning
Example: Market segmentation

Market segmentation: Group customers into different market segments

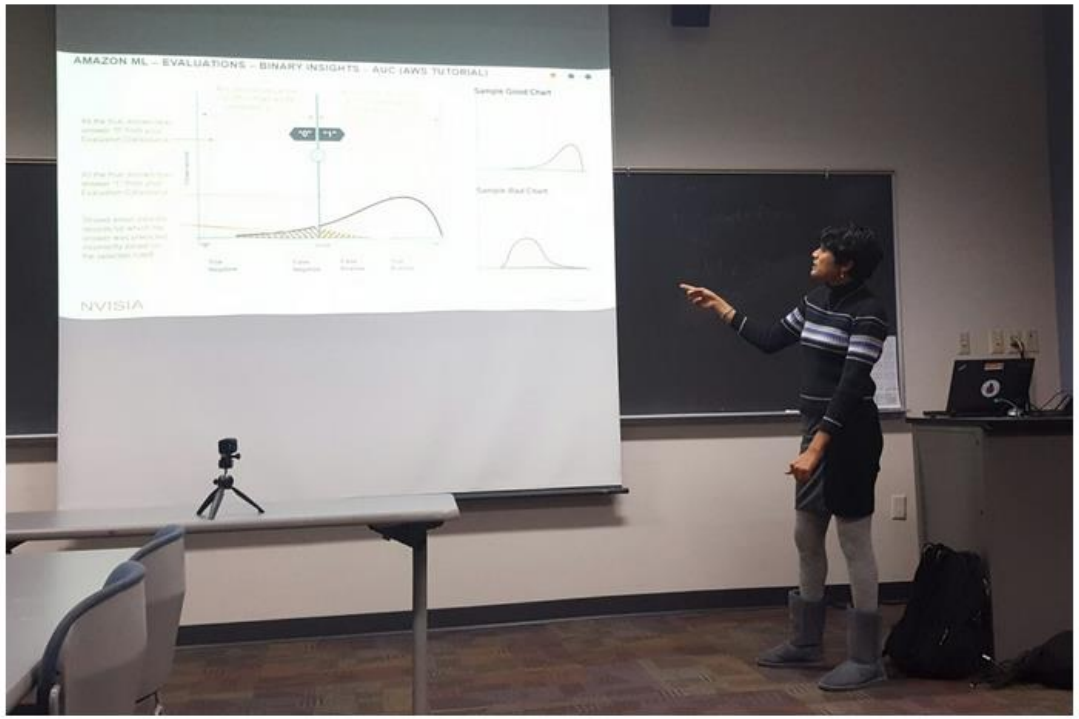


Machine Learning … visual data (4)

Linear Regression (Simplified): Predict house price: Gradient Descent Plot



Predictive Analytics



PREDICTIVE ANALYTICS WITH BIG DATA

Posted on January 10th, 2017

Naveen VK demonstrated how to use historical data and modern tools to learn to predict future trends based on existing big data. Using a dataset from kaggle.com, Naveen

Predictive Analytics … visual data (1)

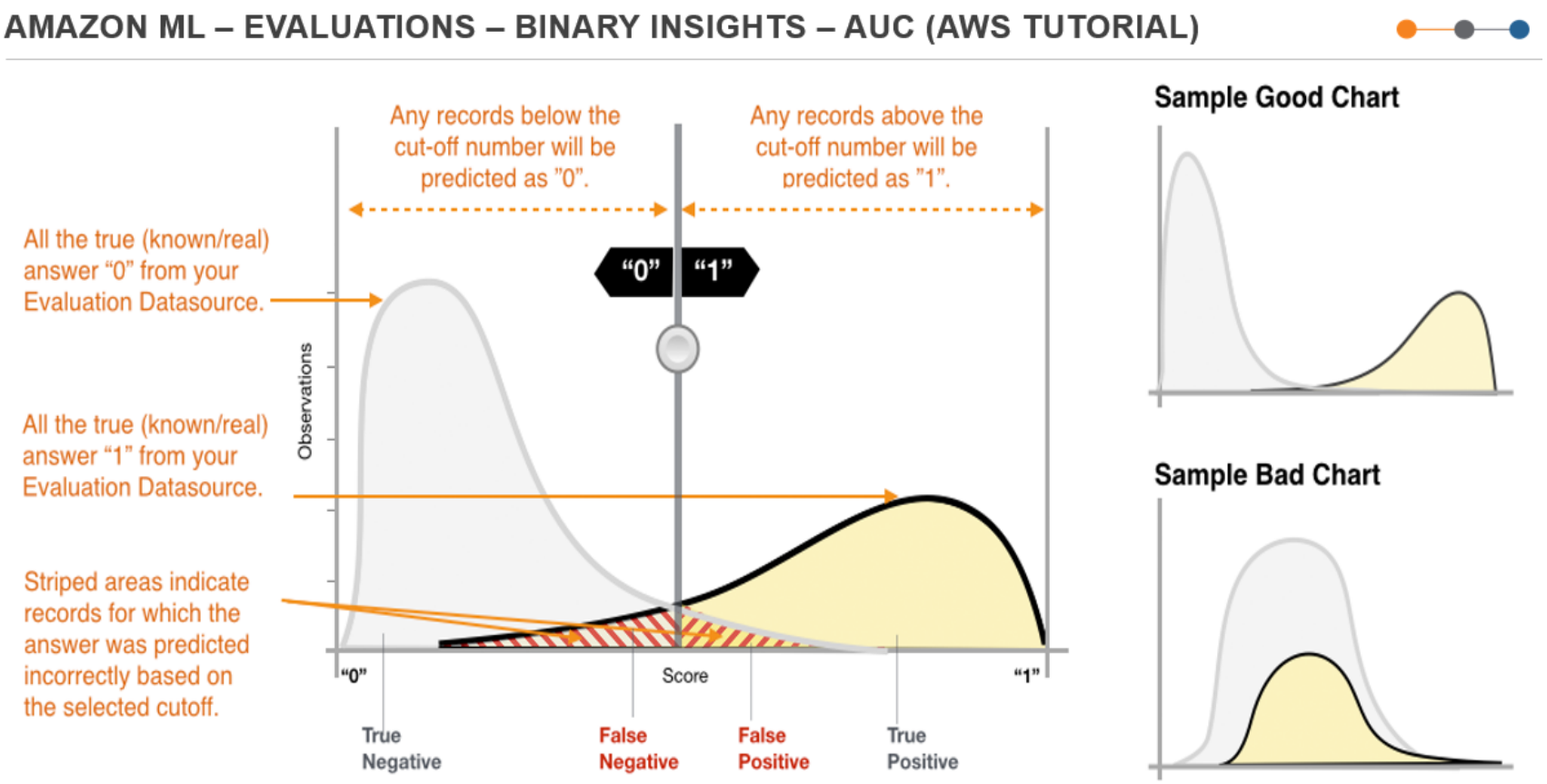
USECASE – TITANIC SURVIVORS

Did this person survive the sinking of the Titanic?

- Titanic Survivors Dataset from kaggle.com
- CSV file
- Hard to code
  - Survival depends on various parameters: age, cabin/class, gender, siblings/spouse, etc.

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450
6	0	3	Moran, Mr. James	male		0	0	330877
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463

Predictive Analytics … visual data (2)



What can Tableau do?

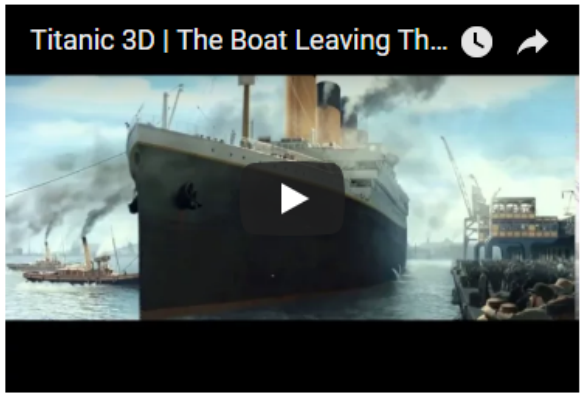
- Can it show Titanic survivor data?
- Can it predict whether a passenger would survive?
- Can it make obvious which parameters strongly influence survival?
- Can it be used for ‘predictive analytics?’
- Can ‘human learning’ via Tableau make ‘machine learning’ less necessary?
- Or is Tableau mostly a fancy front-end/ user interface layer?

Titanic Data Set

A Titanic Probability

Thanks to Kaggle and encyclopedia-titanica for the dataset.

This is the last question of Problem set 5. In this problem you will use real data from the Titanic to calculate conditional probabilities and expectations.



tldr: the ship sinks

On April 15, 1912, the largest passenger liner ever made collided with an iceberg during her maiden voyage. When the Titanic sank it killed 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships. One of the reasons that the shipwreck resulted in such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others.

Update (May/12): We removed commas from the name field in the dataset to make parsing easier.

The titanic.csv file contains data for 887 of the real Titanic passengers. Each row represents one person. The columns describe different attributes about the person including whether they survived (S), their age (A), their passenger-class (C), their sex (G) and the fare they paid (X).

Titanic Dataset

[Quetion12] Write a program in C, C++, Java or Python that **reads the data file** and finds the answers to the following questions:

a. Calculate the conditional probability that a person survives given their sex and passenger-class:

http://web.stanford.edu/class/archive/cs/cs109/cs109.1166/problem12.html

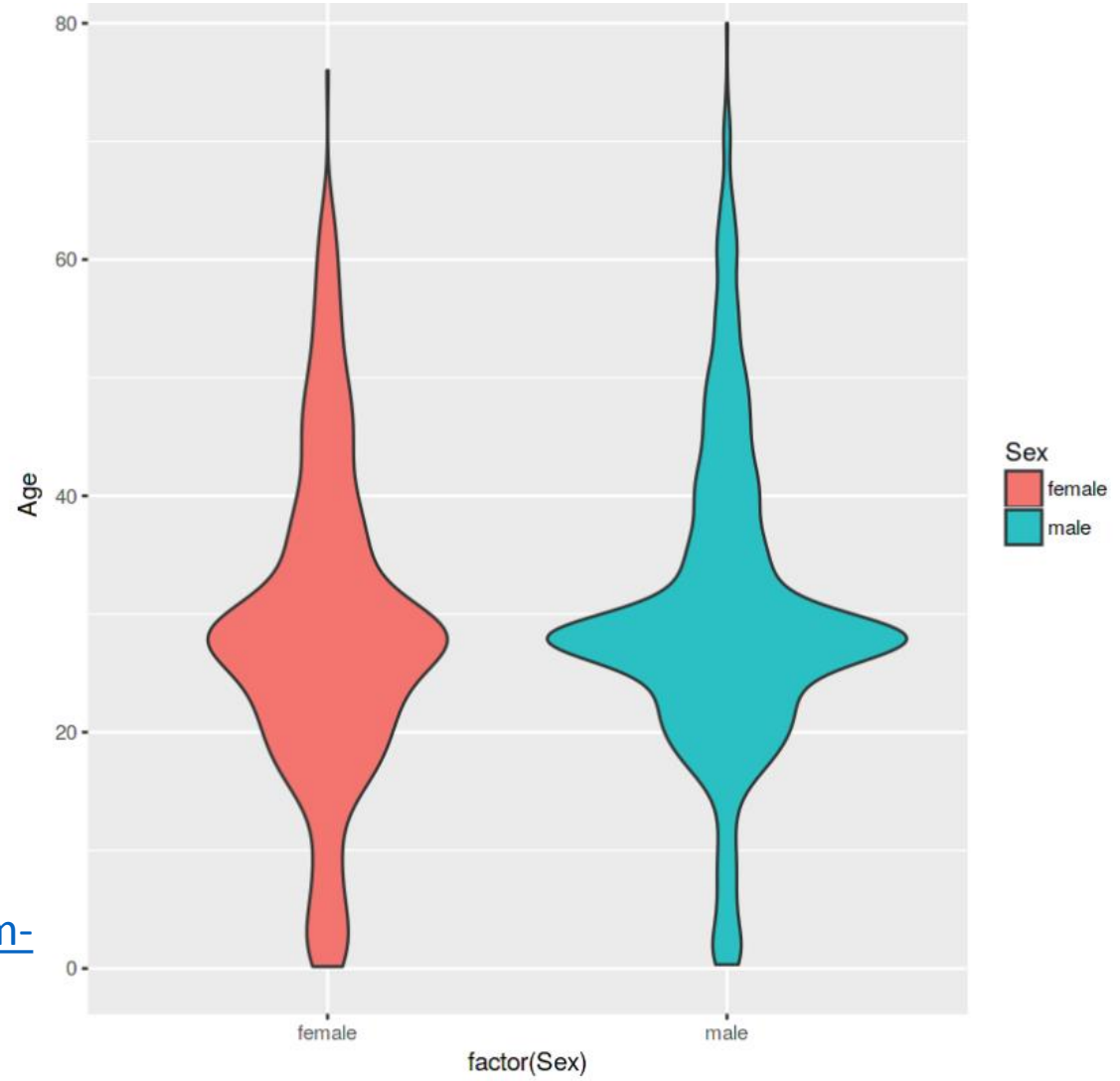
Titanic Data Set (CS homework)

- [Quetion12] Write a program in C, C++, Java or Python that **reads the data file** and finds the answers to the following questions:
- What is the probability that a child who is in third class and is 10 years old or younger survives?
  - Since the number of data points that satisfy the condition is small use the "bayesian" approach and represent your probability as a beta distribution. Calculate a belief distribution for: S= true | A≤10,C=3
  - You can express your answer as a parameterized distribution.

http://web.stanford.edu/class/archive/cs/cs109/cs109.1166/problem12.html

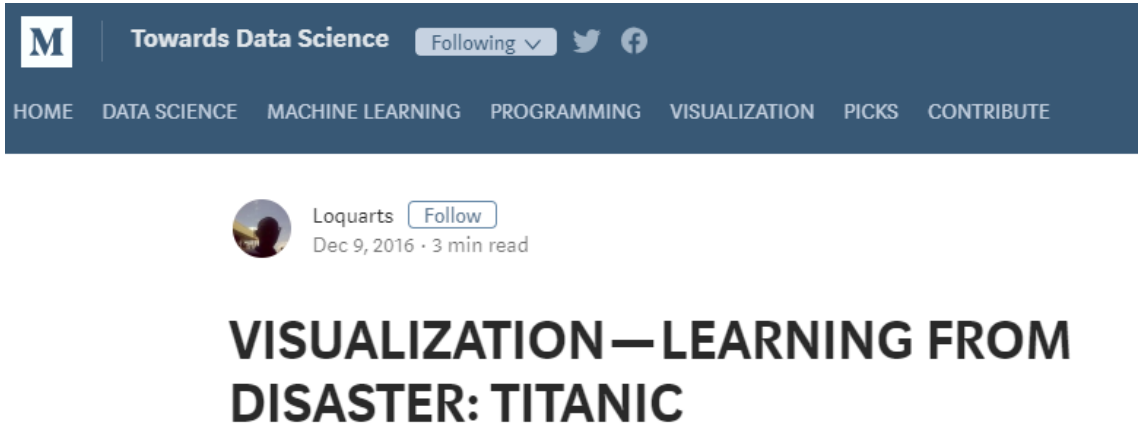
Titanic Data Set (using r)

- Many ways to interpret data
- Using ‘r’ and ‘ggplot2’



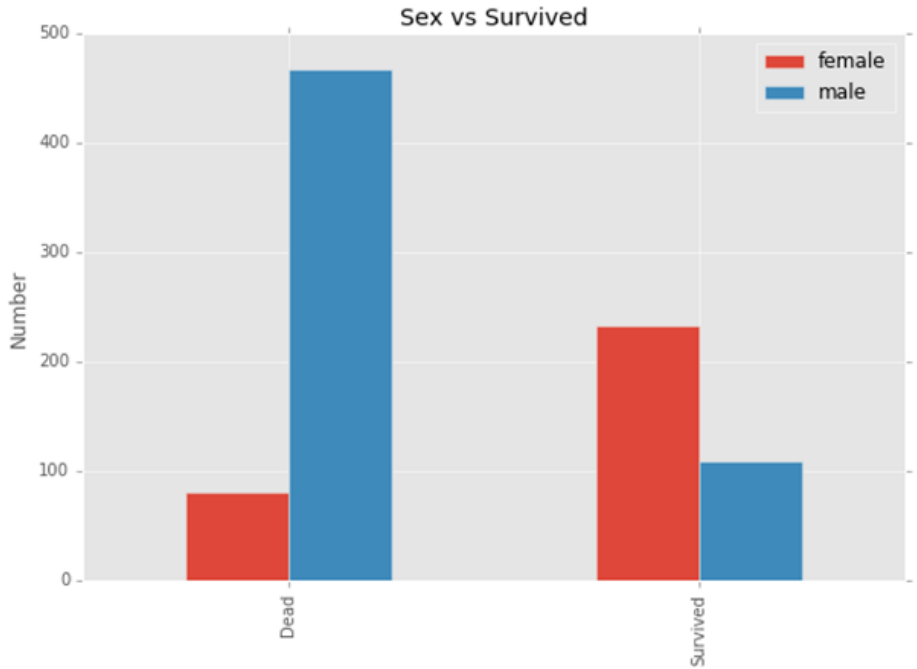
https://towardsdatascience.com/visualization-learning-from-disaster-titanic-42eeb99cdbdc

Titanic Data Set (using Tableau)



https://towardsdatascience.com/visualization-learning-from-disaster-titanic-42eeb99cdbdc

SEX OF PASSENGERS



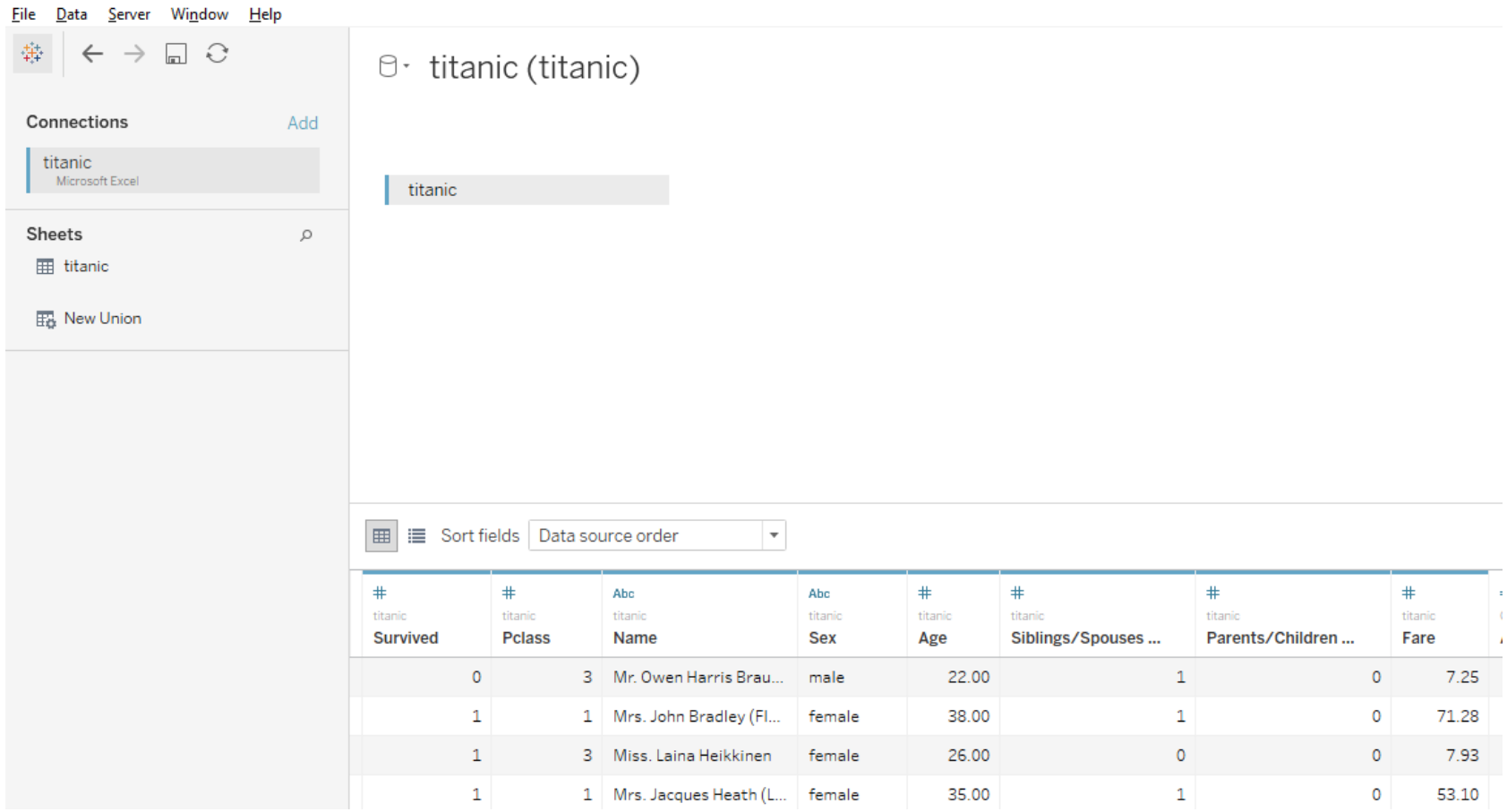
More females survive than males.

Titanic data exploration with Tableau

Download the (clean) data set

	A	B	C	D	E	F	G	H
1	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
2	0	3	Mr. Owen Harris Braund	male	22	1	0	7.25
3	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cumings	female	38	1	0	71.2833
4	1	3	Miss. Laina Heikkinen	female	26	0	0	7.925
5	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35	1	0	53.1
6	0	3	Mr. William Henry Allen	male	35	0	0	8.05
7	0	3	Mr. James Moran	male	27	0	0	8.4583
8	0	1	Mr. Timothy J McCarthy	male	54	0	0	51.8625
9	0	3	Master. Gosta Leonard Palsson	male	2	3	1	21.075
10	1	3	Mrs. Oscar W (Elisabeth Vilhelmina Berg) Johnson	female	27	0	2	11.1333

Create a data source



Begin data exploration

age and gender

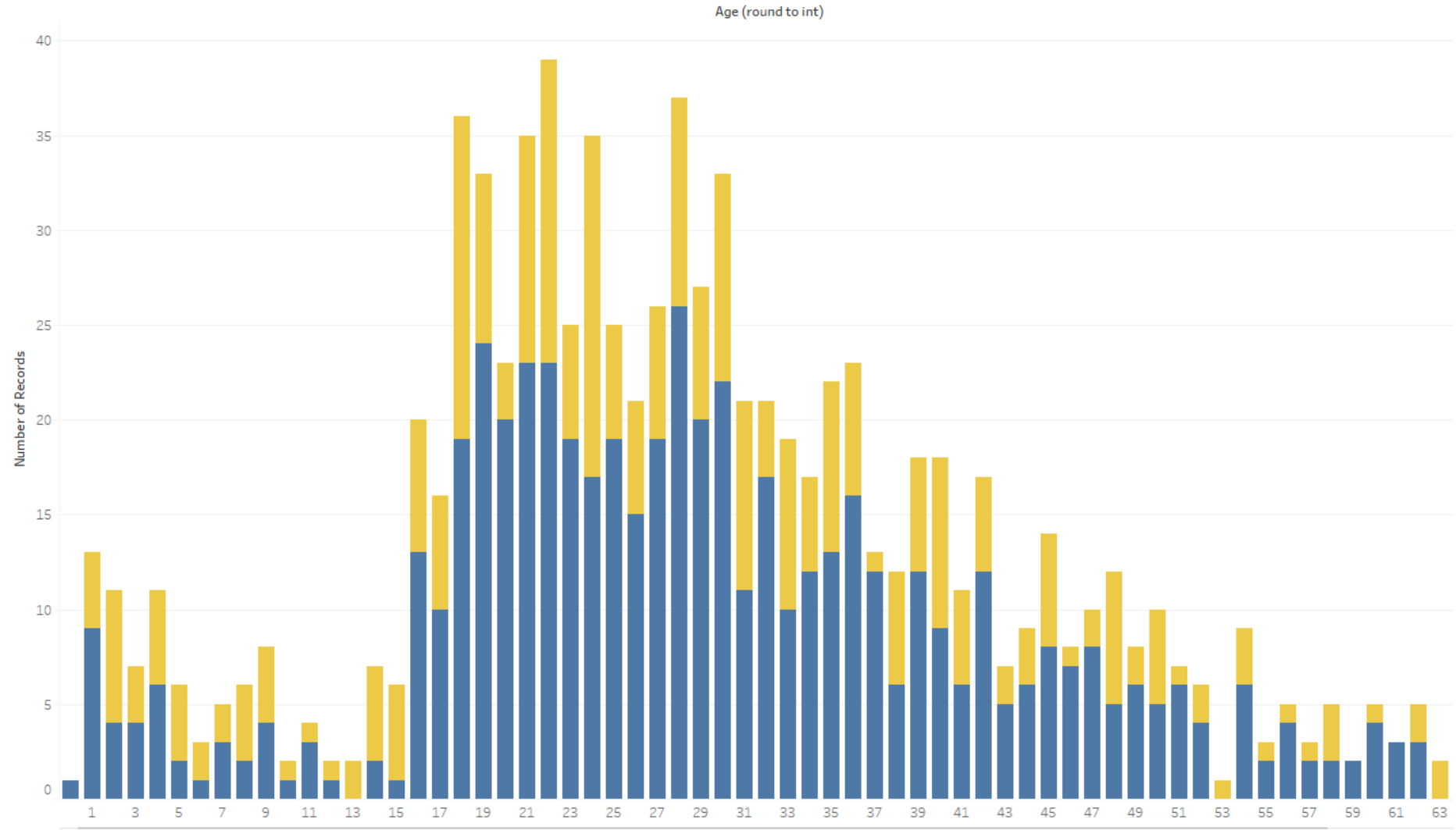
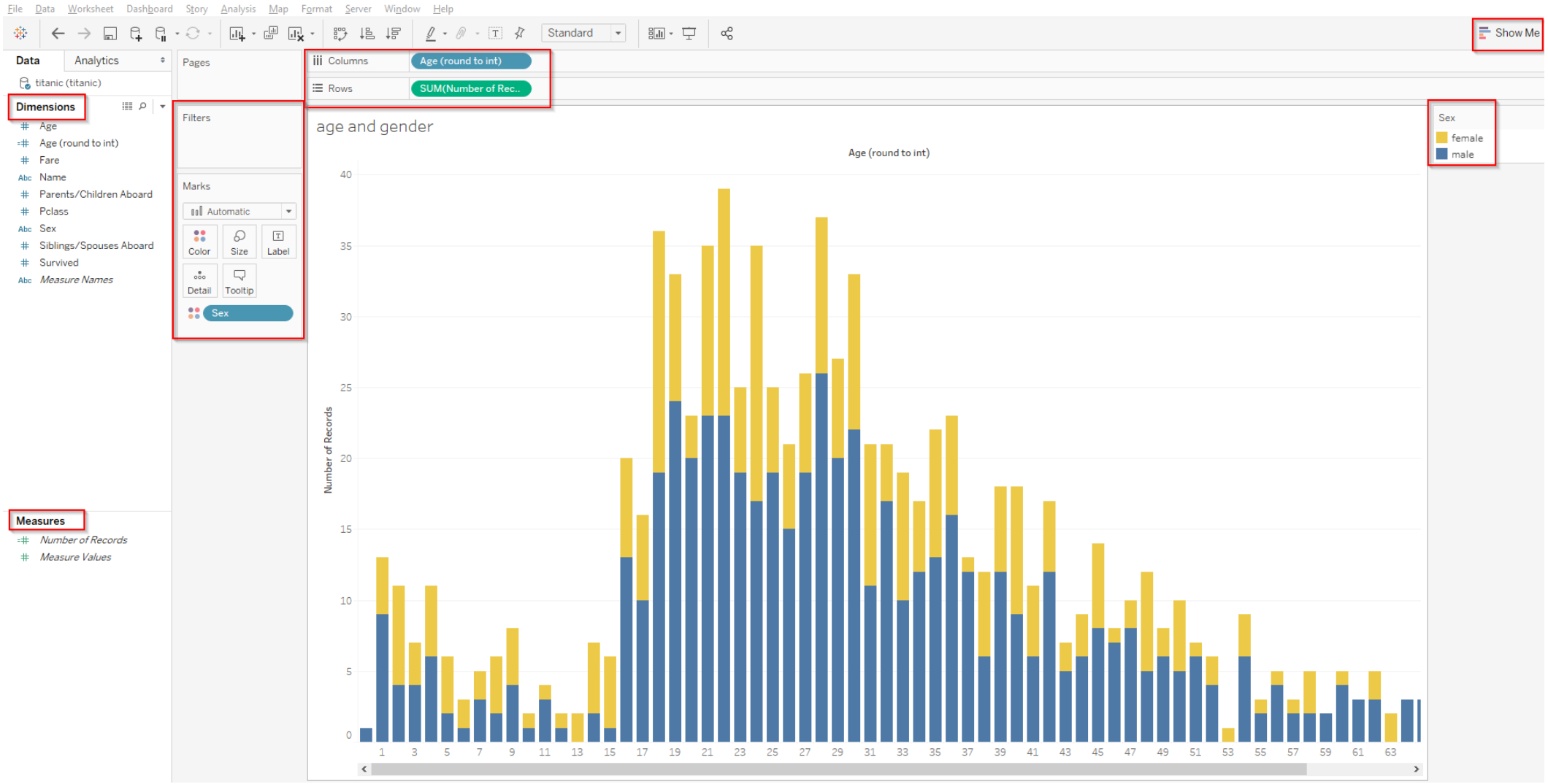
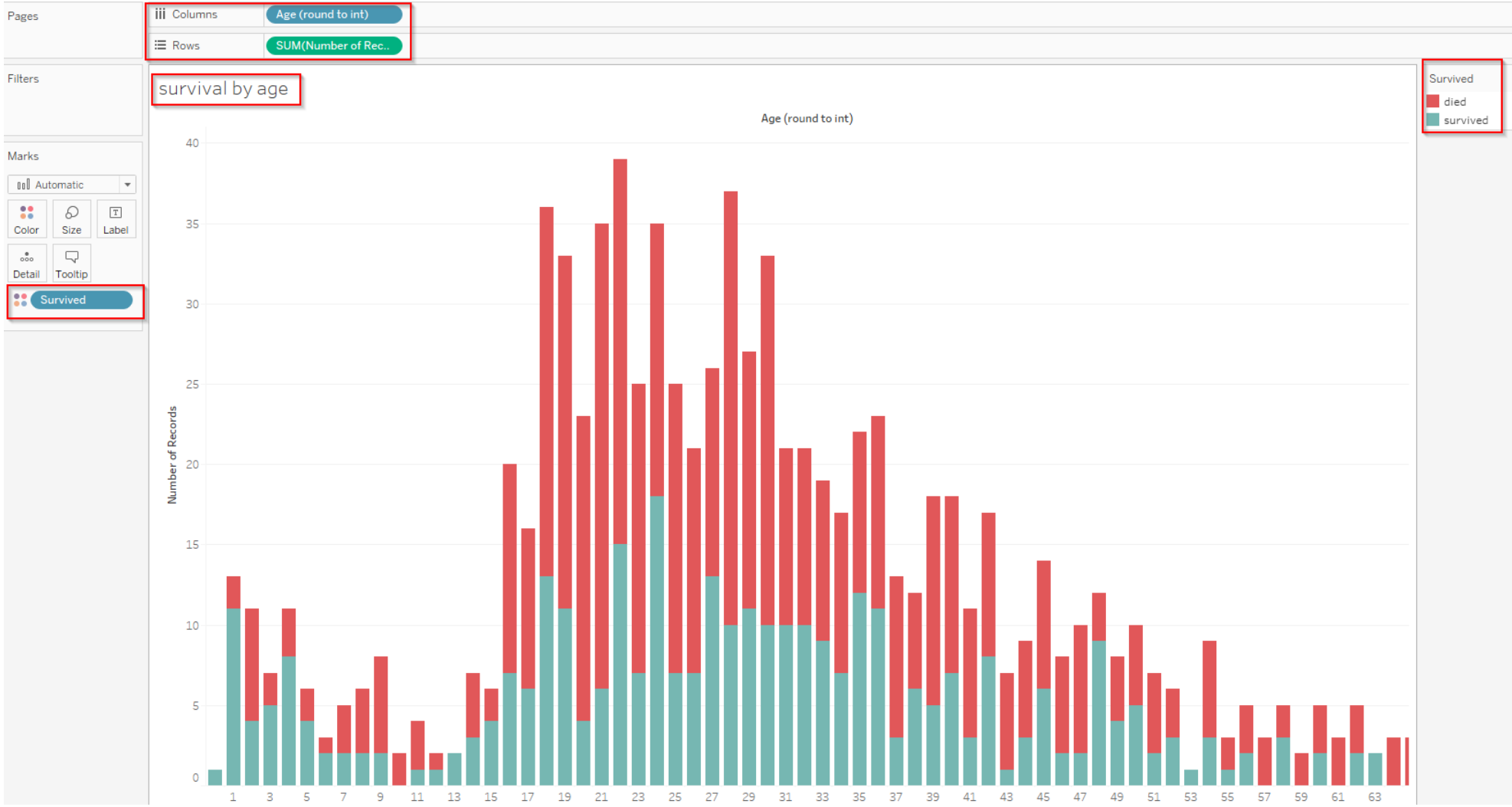


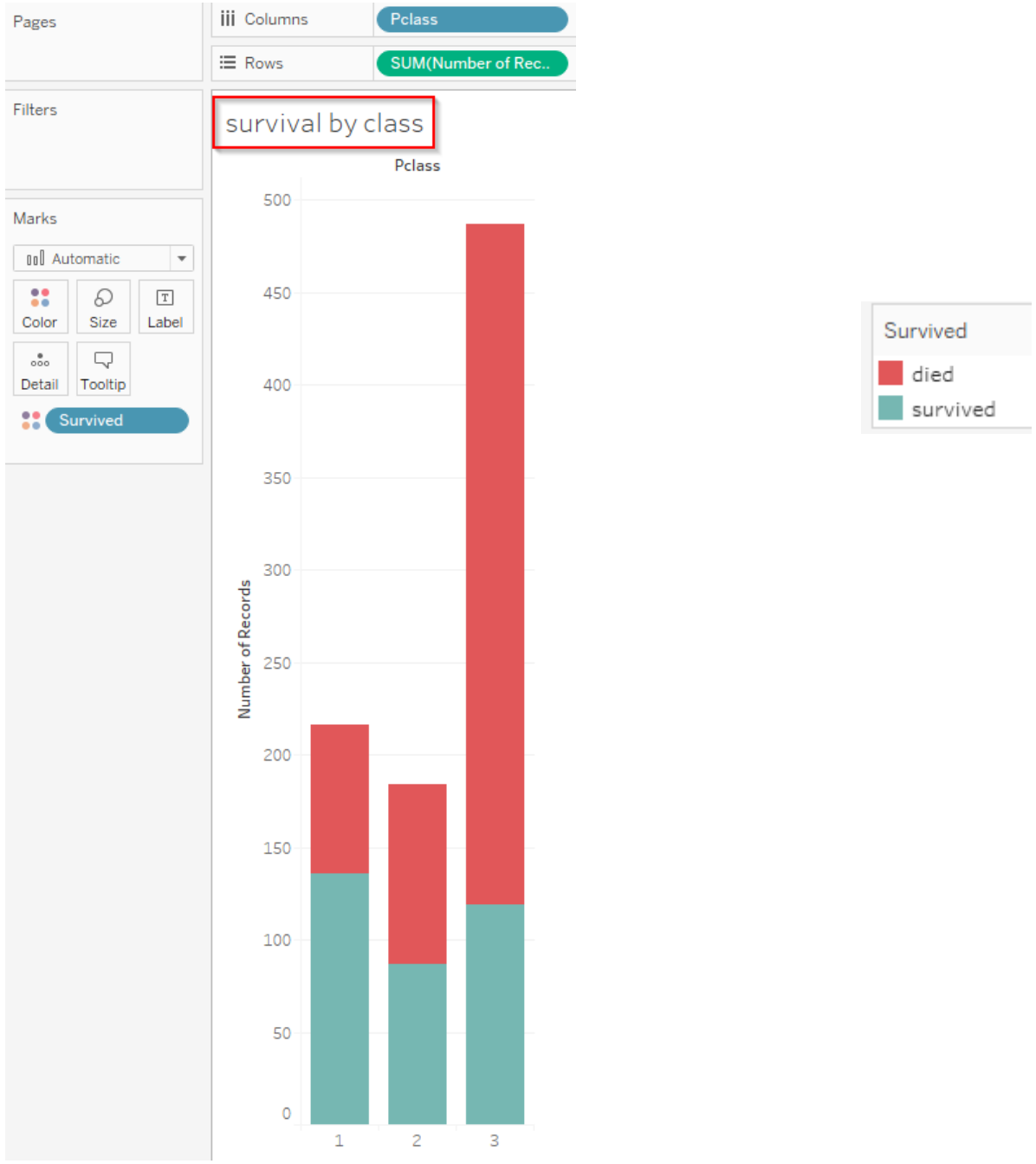
Tableau display options



Survival by age

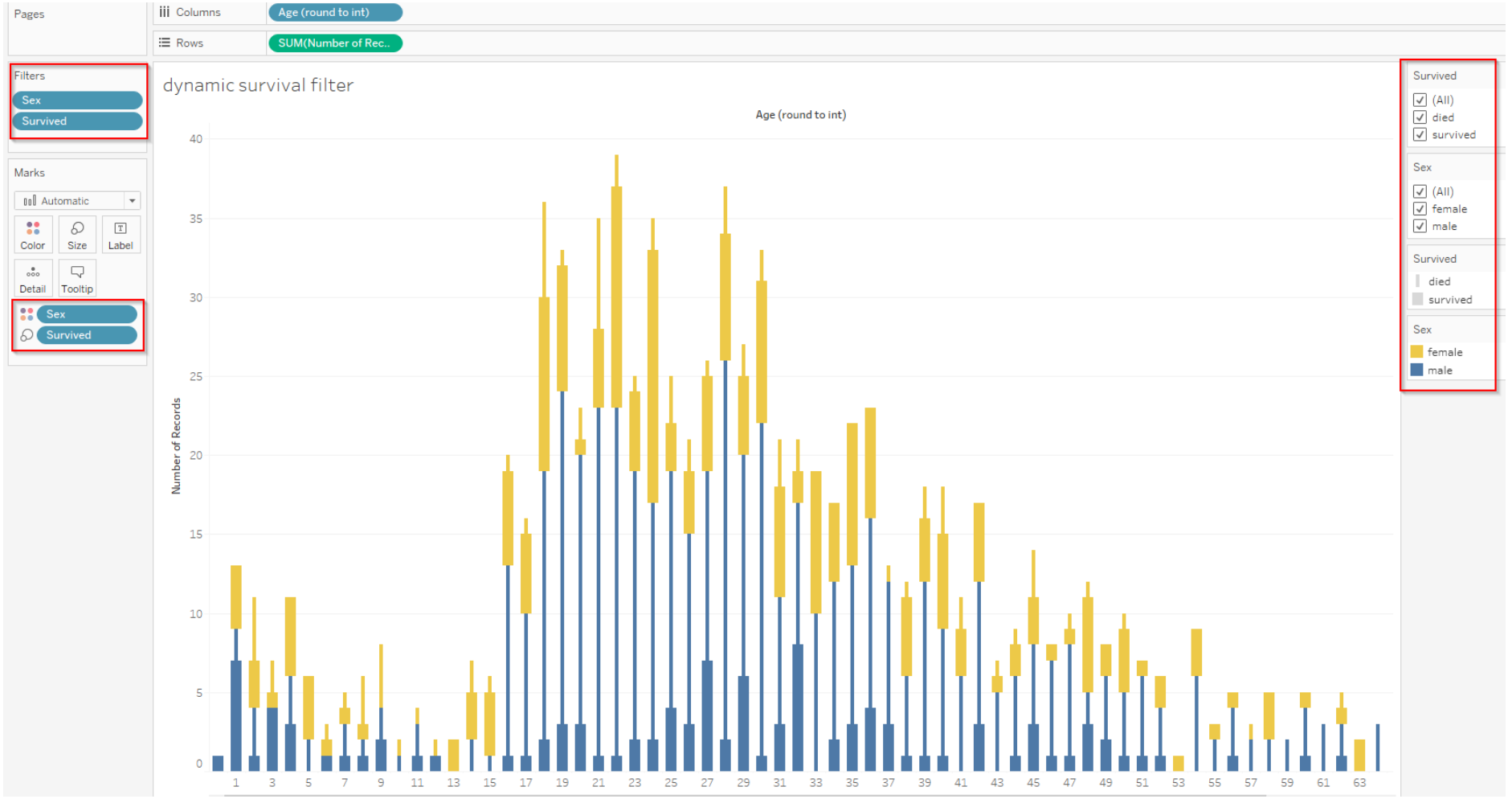


Survival by class



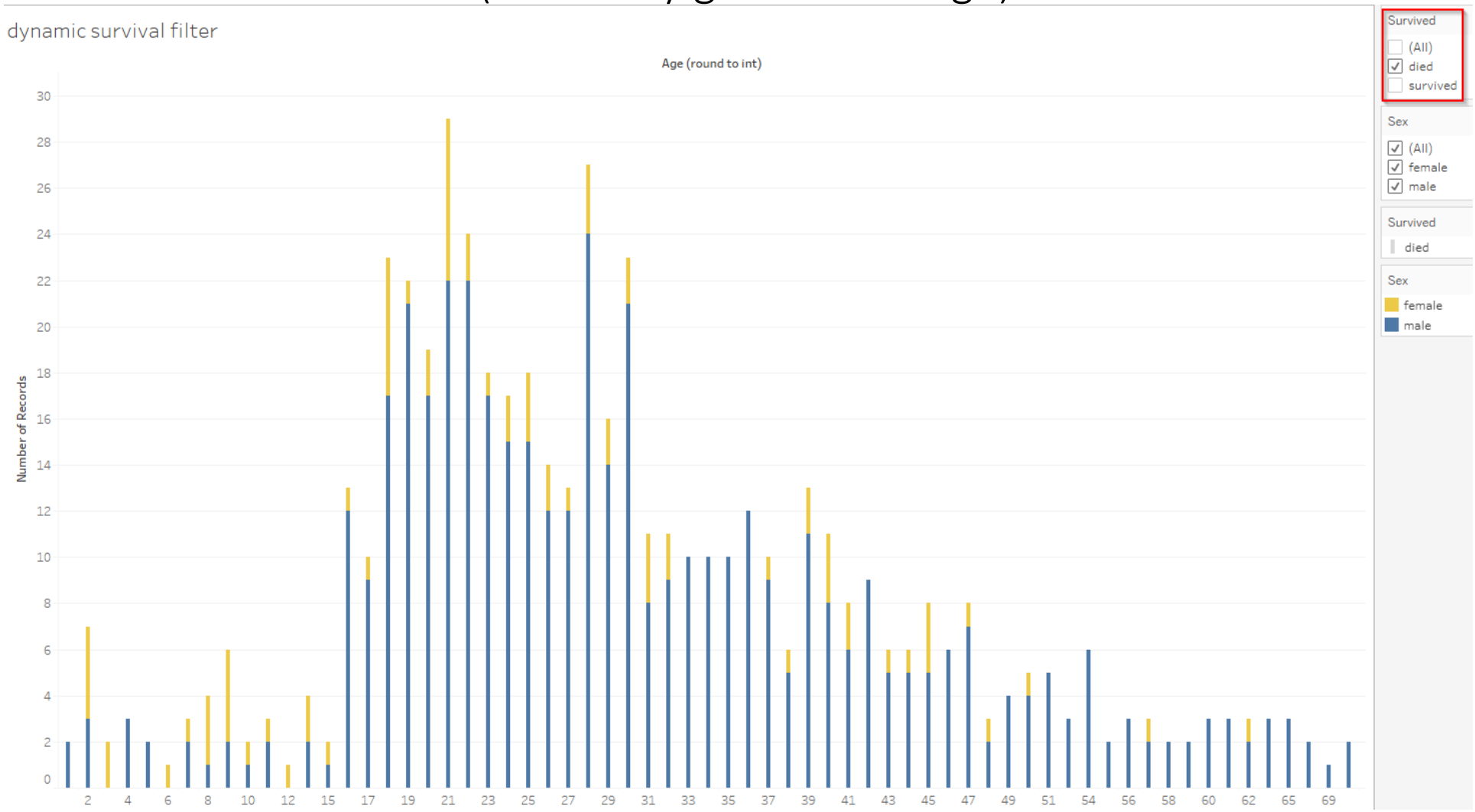
Dynamic filters

(user can ask questions and answer them)



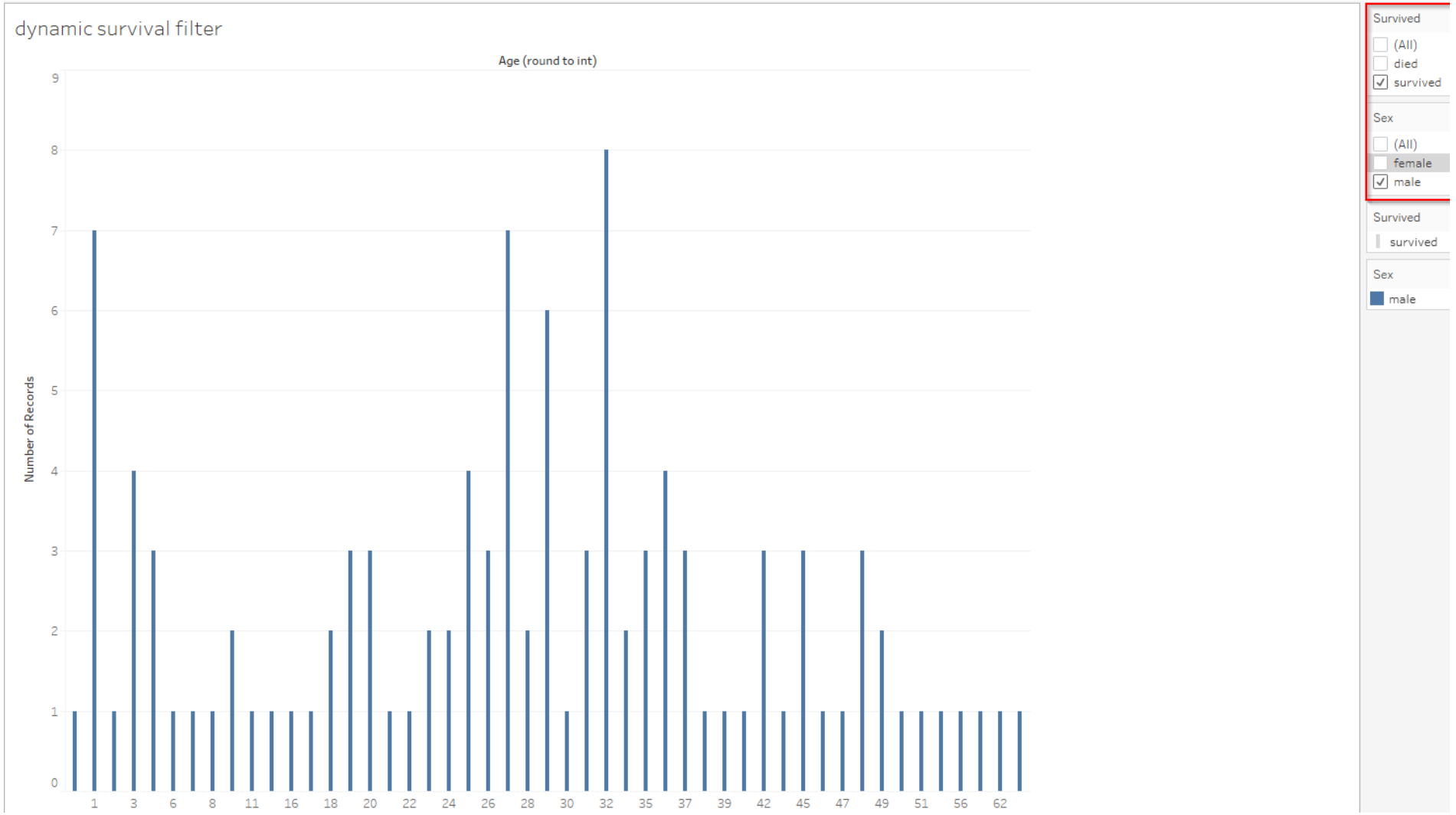
Don’t make ugly, complex visuals!!

(deaths by gender and age)

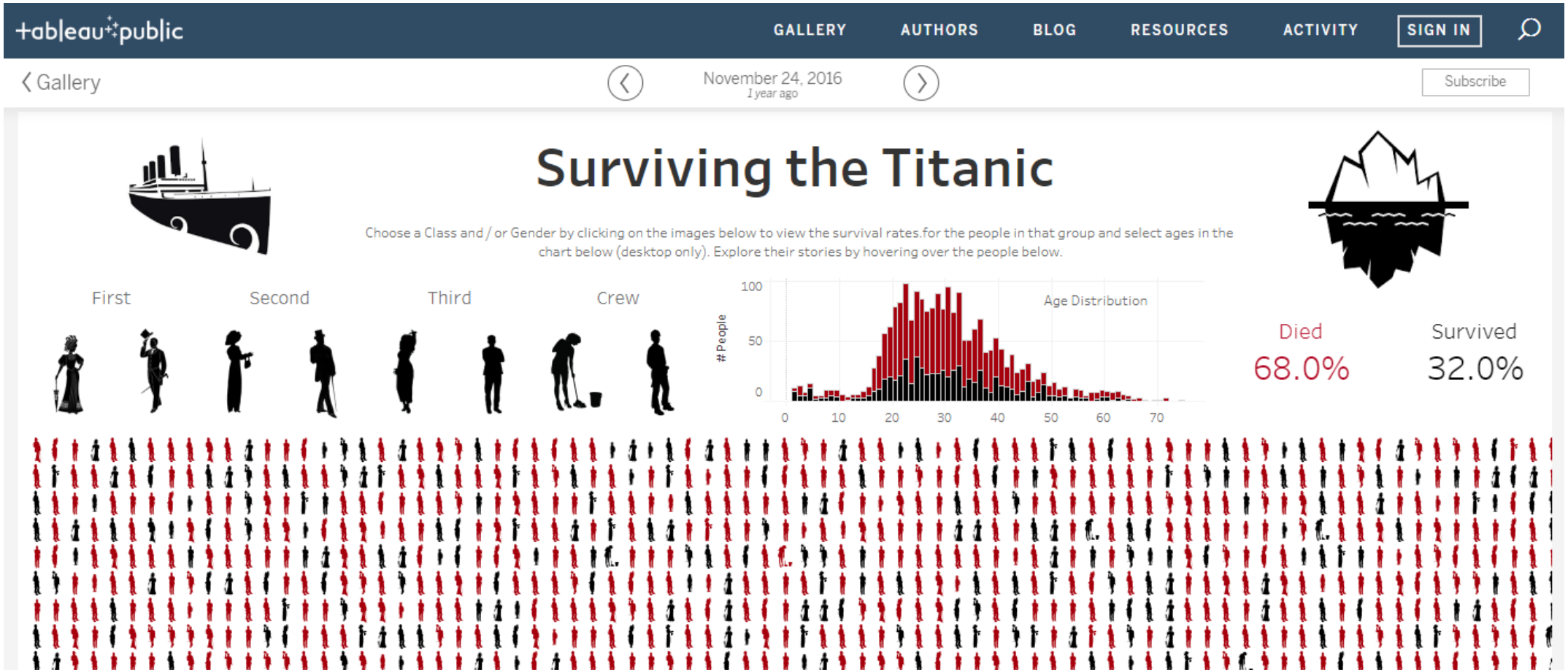


Don’t make ugly, complex visuals!!

(survival by age for males)

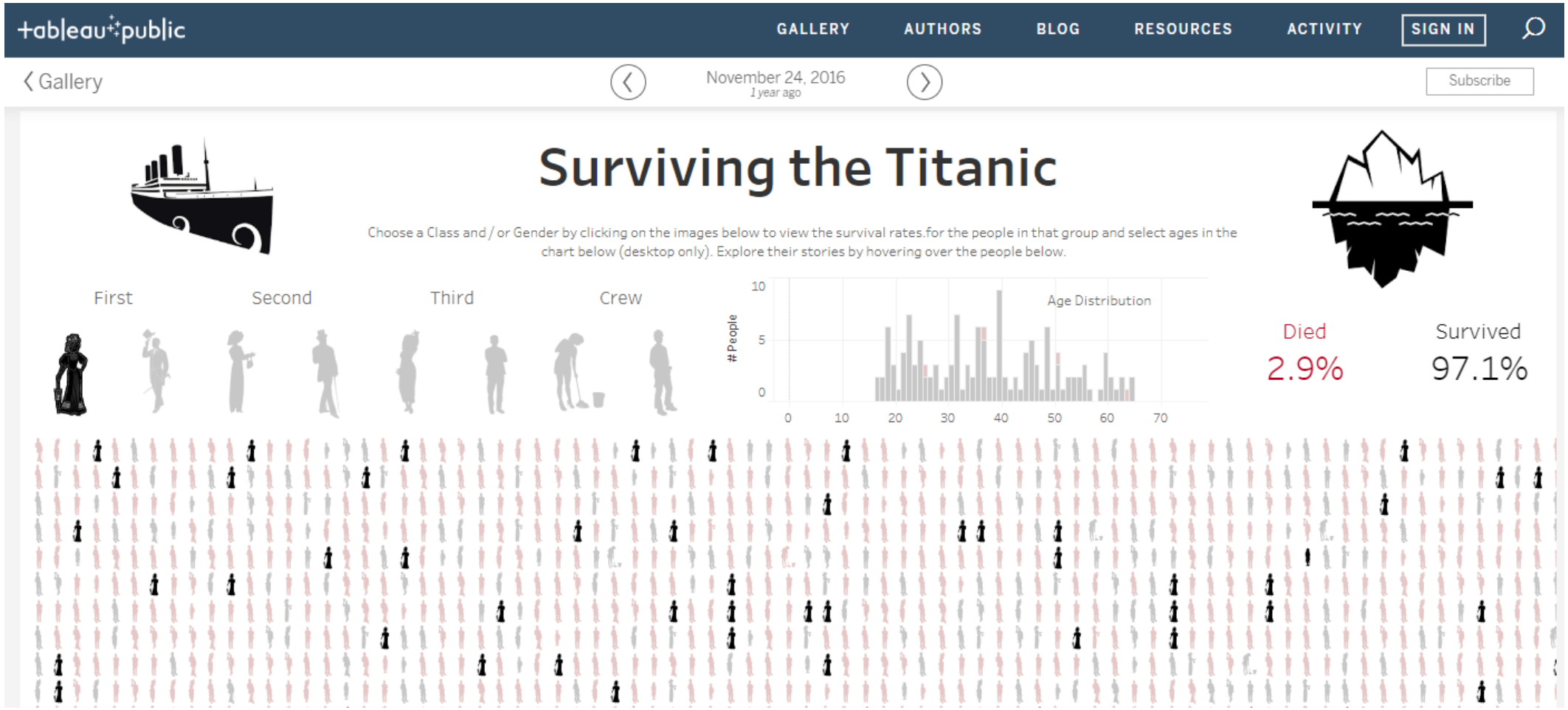


Look to the professional for true art!



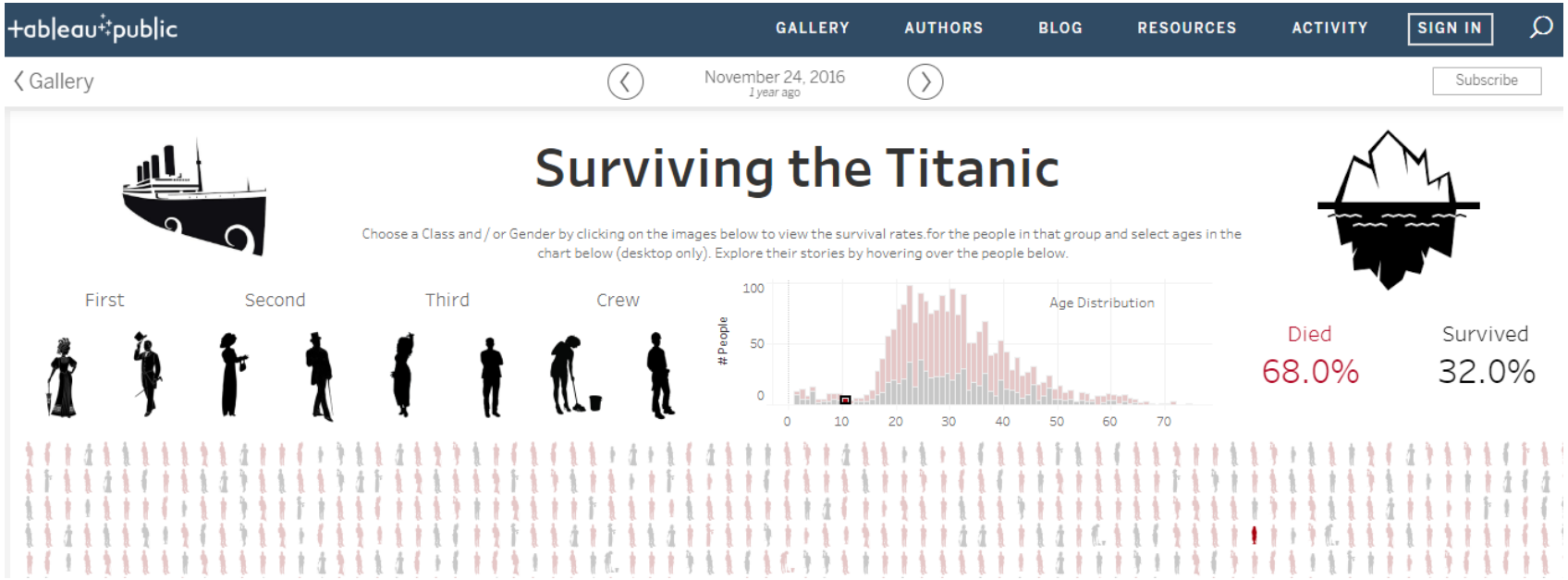
https://public.tableau.com/en-us/s/gallery/surviving-titanic

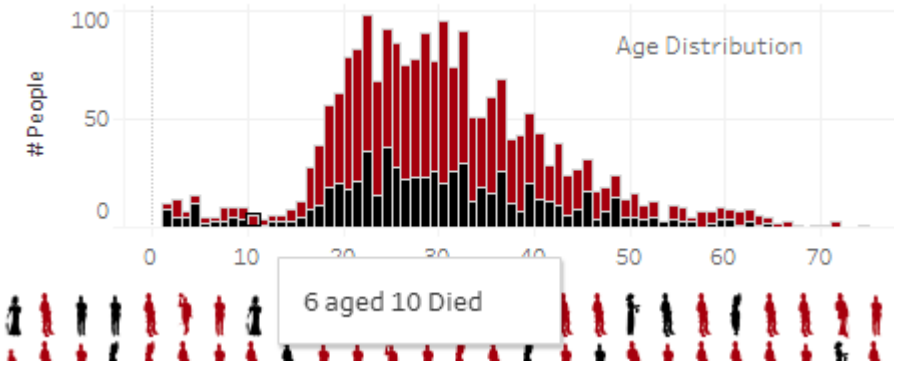
Female first class data



https://public.tableau.com/en-us/s/gallery/surviving-titanic

Male 10-year old data





https://public.tableau.com/en-us/s/gallery/surviving-titanic

What can Tableau do?

- Can it show Titanic survivor data?
- Can it predict whether a passenger would survive?
- Can it make obvious which parameters strongly influence survival?
- Can it be used for ‘predictive analytics?’
- Can ‘human learning’ via Tableau make ‘machine learning’ less necessary?
- Or is Tableau mostly a fancy front-end/ user interface layer?

Karen’s conclusions…

- Tableau can be good for data exploration and insights
- Tableau can increase the ‘data literacy’ of non-technical people
- Tableau will not replace the need for:
  - Predictive analytics
  - Machine learning
  - Big data use

Melissa's conclusions…

- Tableau is a communications tool
  - Quick data understanding (data exploration)
  - Visually beautiful
    - Aesthetics matter
    - Low threshold to data literacy
    - Makes data accessible to non-technical audience

- Remember to communicate what you find in the math and data!!!
- We are accountable to management and customers!
- Data visualizations ideally direct, focus and support management decisions and future behaviors.

Up next: real world Tableau …