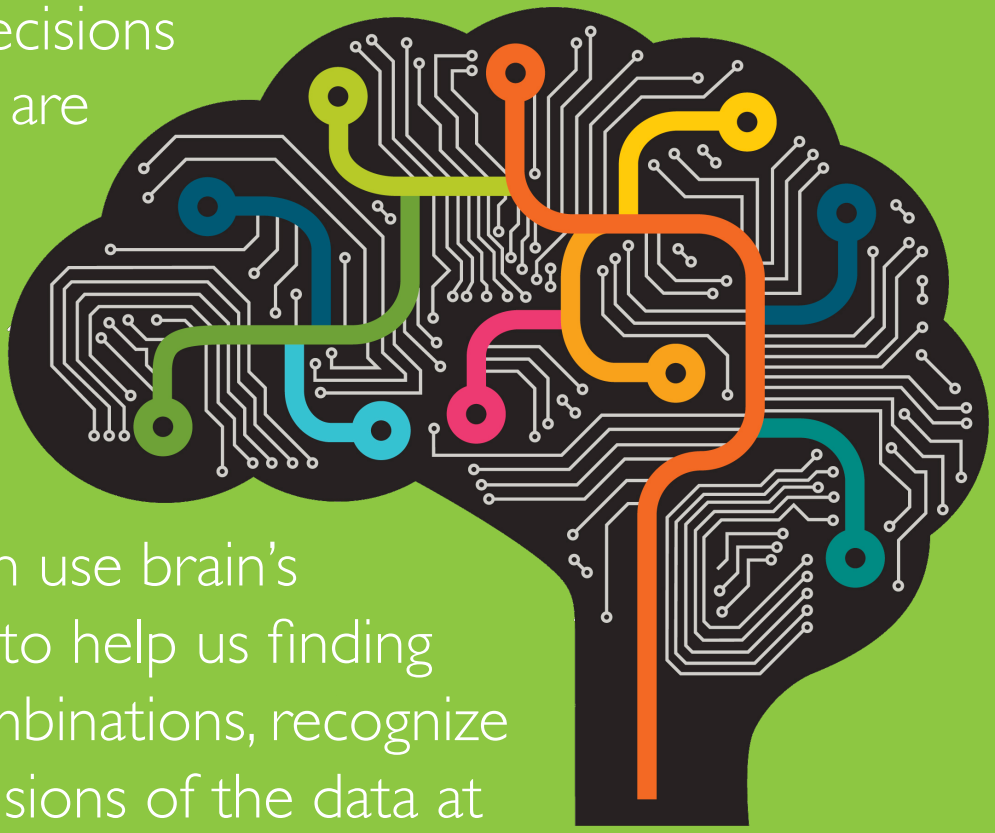
Visual storytelling and data visualization in numerical simulations

Joel Guerrero
University of Genova + Wolf Dynamics

Giovanni Bailardi & Haileyesus Kifle
DLTM La Spezia

The human brain is hardwired for visual processing

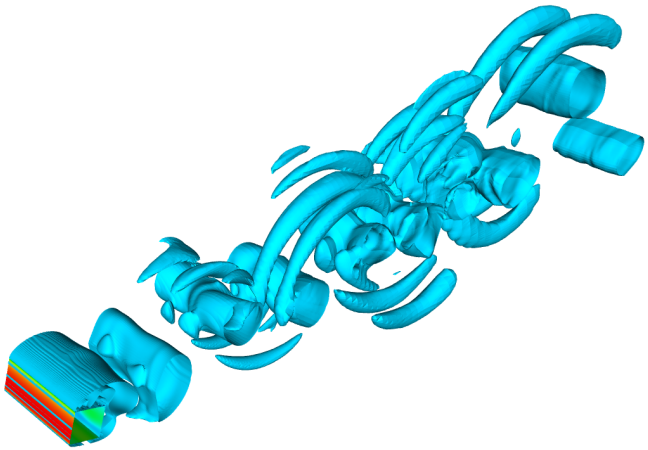
- We have evolved to take snap decisions based on what we perceive. We are really good at recognizing shapes and patterns.
- However, when it comes to crunching numbers or reading fast, we are not that good.
- When working with data, we can use brain's amazing visual processing power to help us finding new insight, explore different combinations, recognize patterns and make informed decisions of the data at a glance.

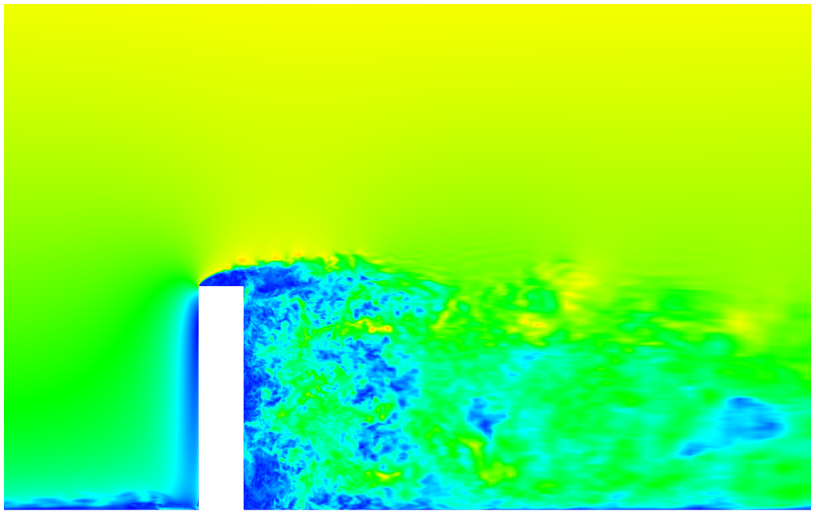


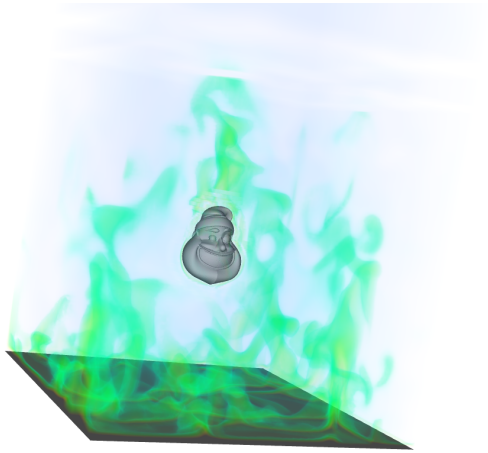
“The greatest value of a picture is when it forces us to notice what we never expected to see.”

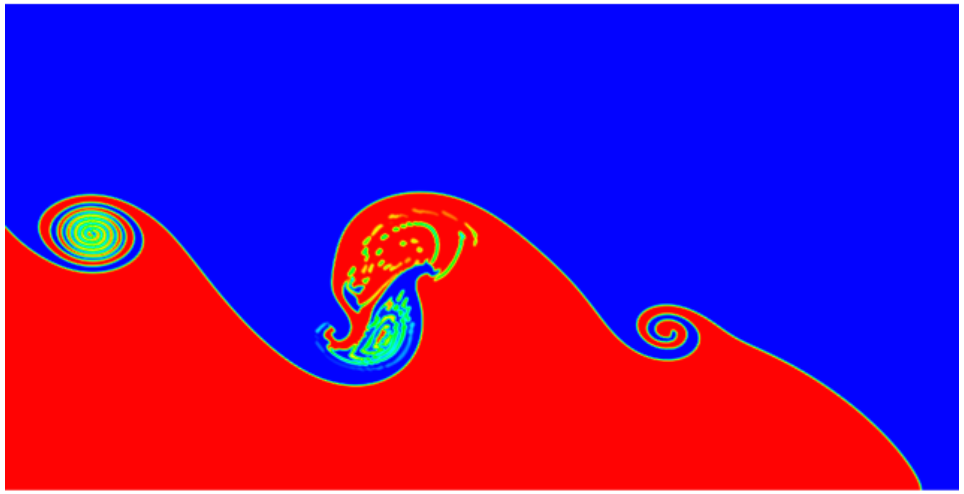
— John Tukey

This presentation is NOT about colorful fluid dynamics (CFD)









This presentation is NOT about colorful fluid dynamics (CFD)

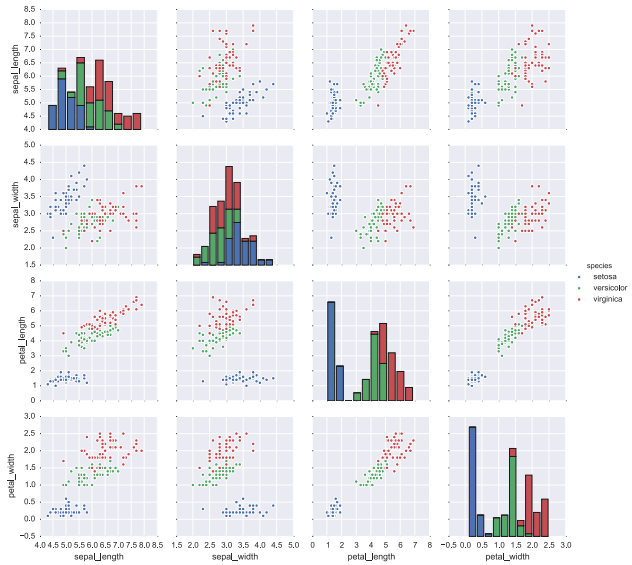
- However, to gather some of the data I ran many numerical simulations. In particular, design space exploration and design optimization studies.
- And believe me, I obtained a lot of data in a relatively short amount of time.
- CFD is not anymore about submitting a few simulations and waiting long times.

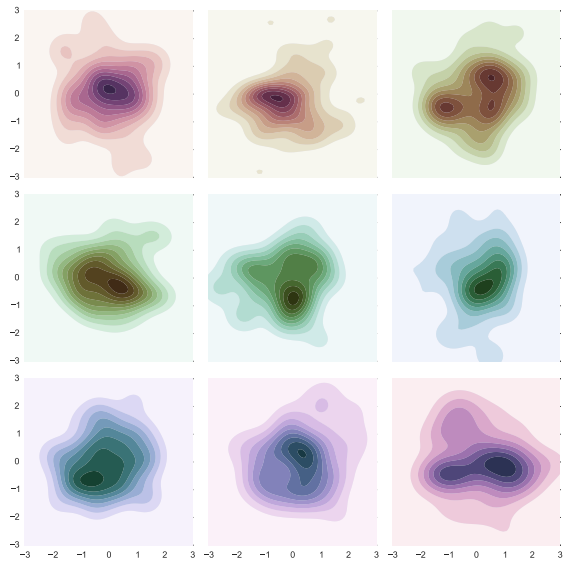
This presentation is NOT about colorful fluid dynamics (CFD)

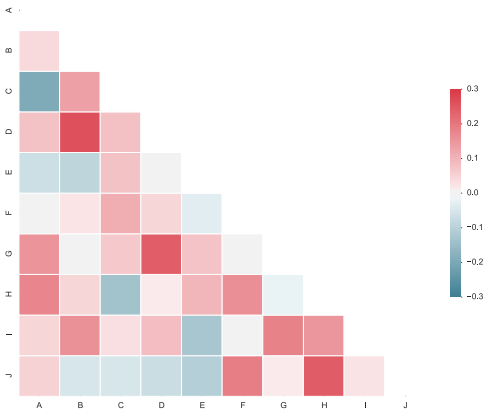
- Thanks to data analitics (DA) and exploratory data analysis* (EDA), I was able to turn all the quantitative information into valuable insight.
- It also helped me in understanding multivariate data and interpreting the Pareto front obtained from multi-objective optimization studies.
- It also opened a new door. The door to interactive manipulation and cross-filtering of the data.

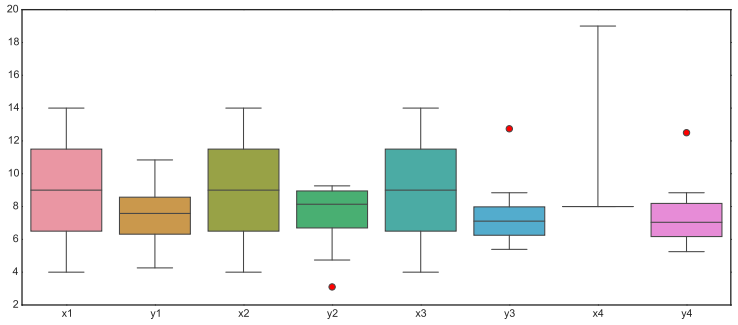
* A fancy term for data visualization. EDA is an approach to analyzing data sets to summarize their main characteristics, often with visual methods.

This presentation is about charts, plots and visual encoders

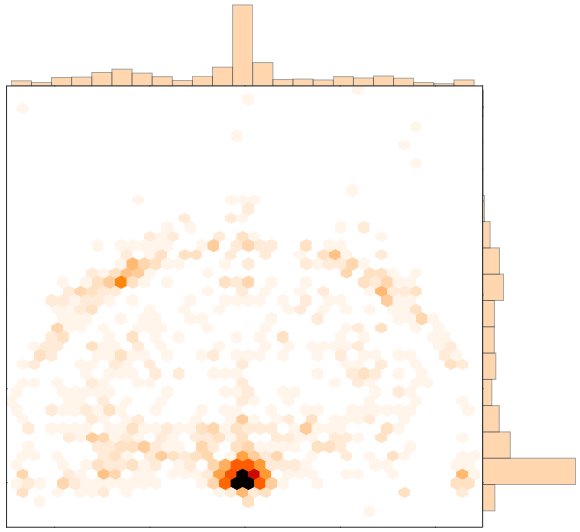


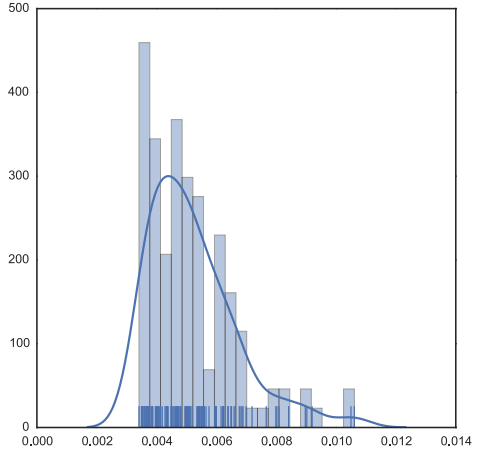


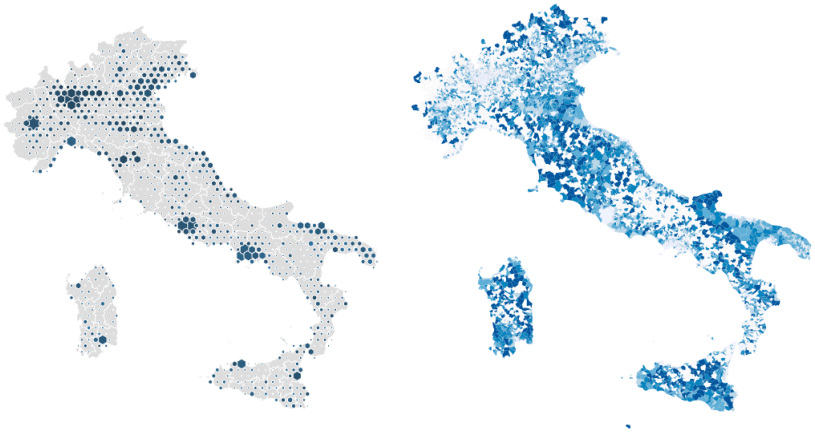


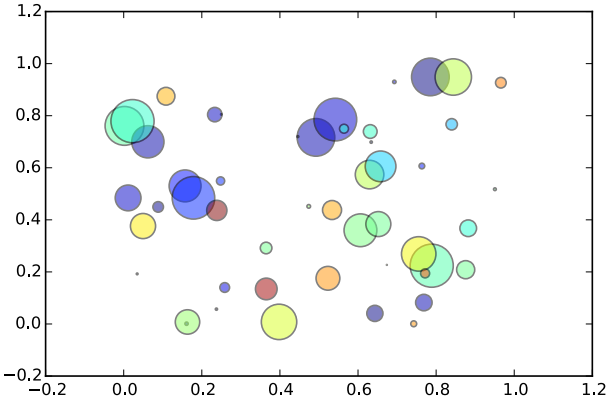


This presentation is about charts, plots and visual encoders

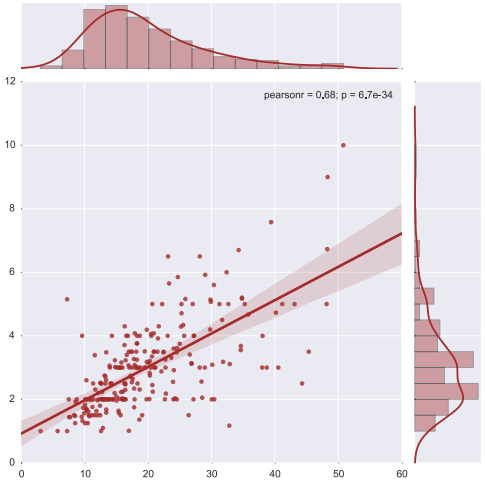


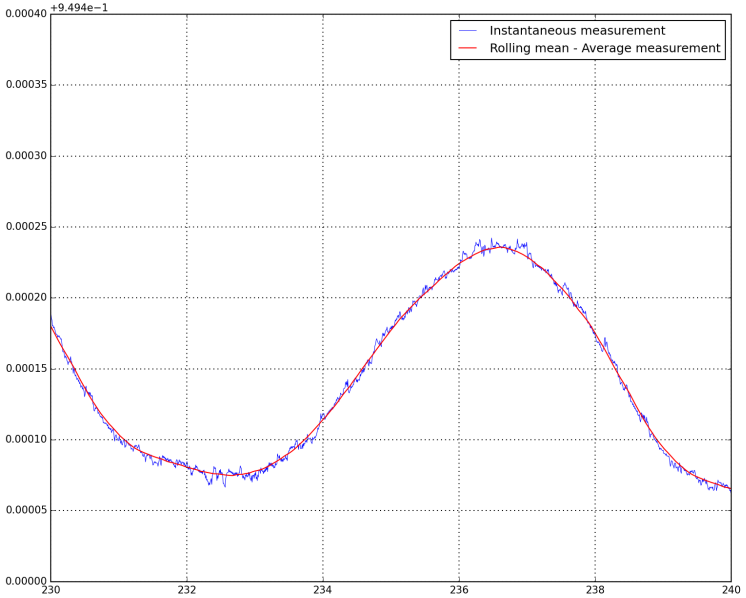


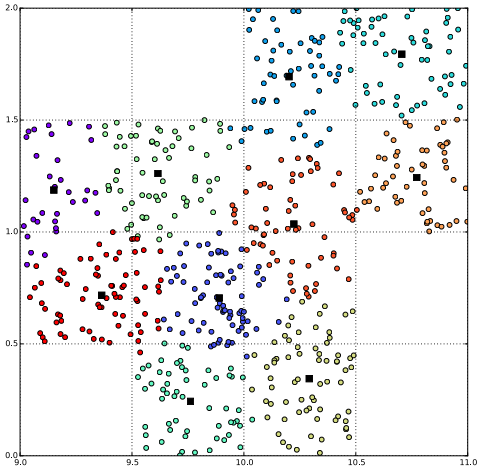


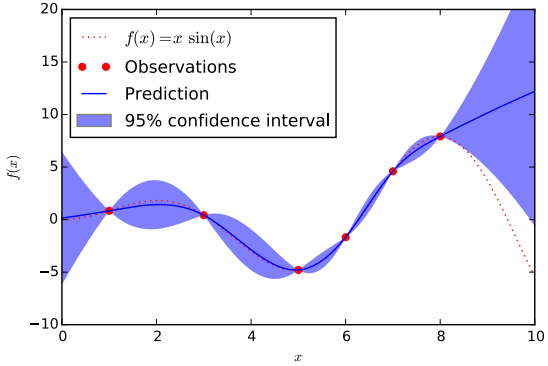


This presentation is about charts, plots and visual encoders



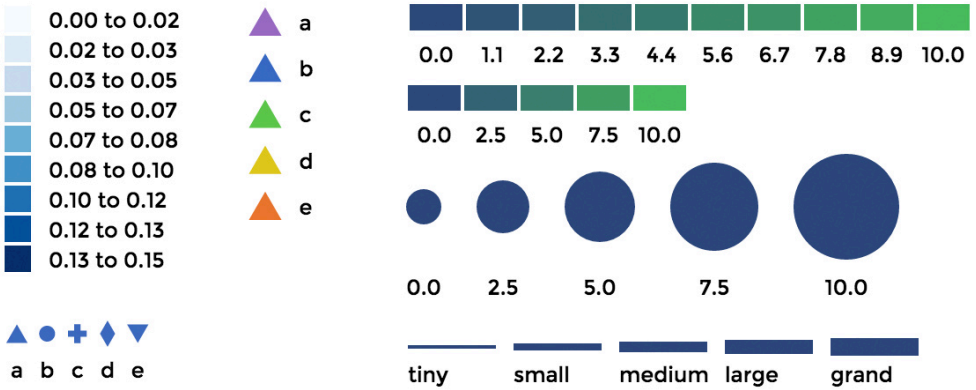






This presentation is about charts, plots and visual encoders

- The key to effectively exposing meaningful patterns in data comes down to thoughtful visual encoding.
- Incremental changes in aesthetics should reflect and be perceived as proportional and meaningful changes in data.

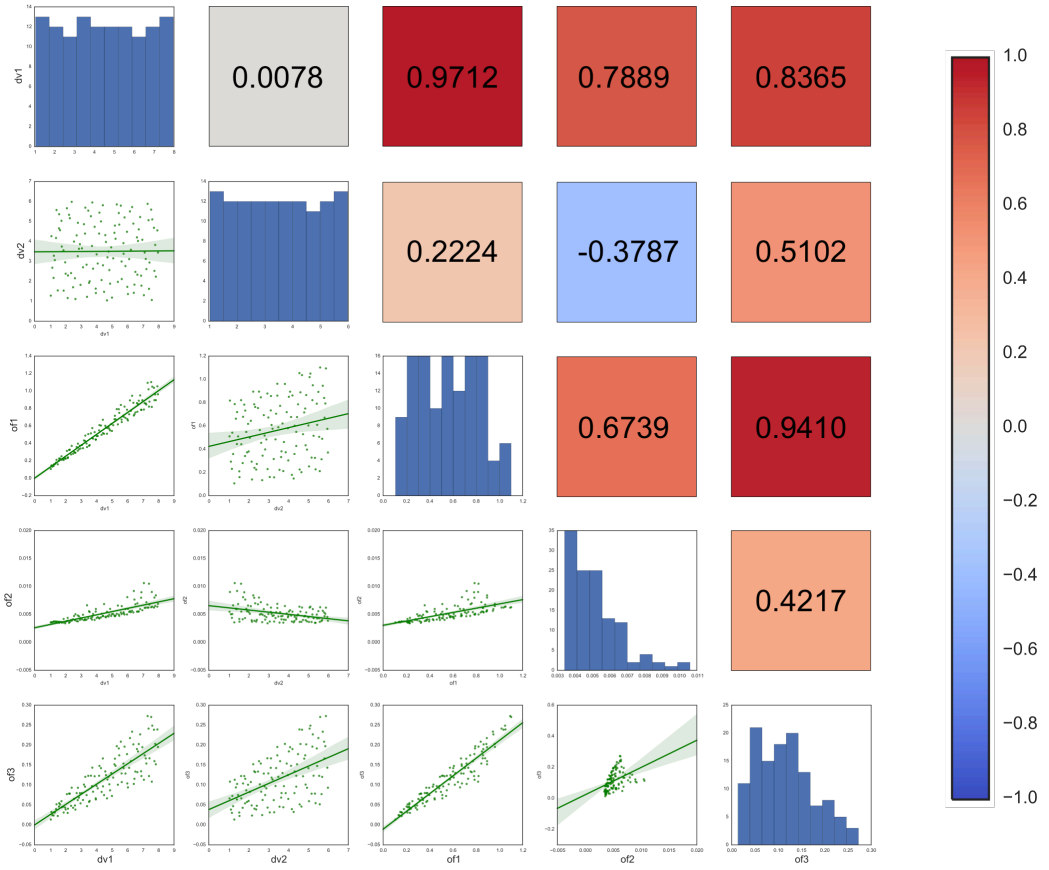


D3 Legend by Susie Lu
http://www.susielu.com/

The what of data visualization and visual storytelling

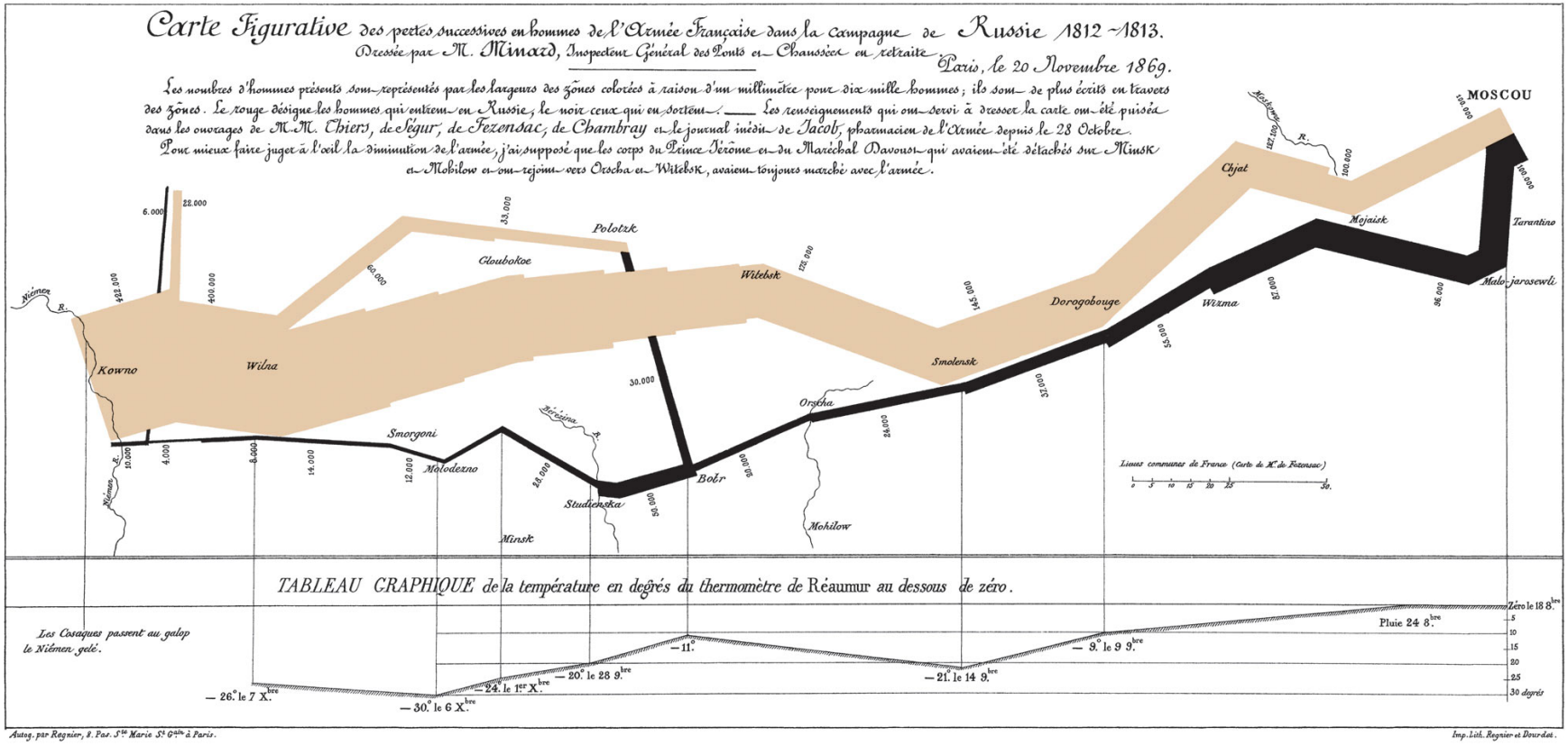
What is data visualization?

- Data visualization is the presentation of data in a pictorial or graphical format in order to amplify cognition.



What is visual storytelling?

- Communication of a story or known information through visual components.



The why of data visualization and visual storytelling

Why data visualization and visual storytelling?

- Patterns, trends, correlations and anomalies that might go undetected in raw data can be exposed and recognized easily when visualizing it.
- Turn data into valuable insights and make informed decisions.
- Spur new questions and prompt skepticism.
- Explore more combinations, interactive manipulation and cross-filtering of data.
- Communicating data in an effective way to a general audience.
- Working in multidisciplinary groups.
- Because we have raw data.

Raw data

Sat Nov 14 21:39:05 CET 2015	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58	Left
Sat Nov 14 21:39:10 CET 2015	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76	Left
Sat Nov 14 21:39:15 CET 2015	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71	Right
Sat Nov 14 21:39:20 CET 2015	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84	Top
Sat Nov 14 21:39:25 CET 2015	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47	Top
Sat Nov 14 21:39:30 CET 2015	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04	Bottom
Sat Nov 14 21:39:35 CET 2015	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25	Bottom
Sat Nov 14 21:39:40 CET 2015	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50	Right
Sat Nov 14 21:39:45 CET 2015	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56	Up
Sat Nov 14 21:39:50 CET 2015	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91	Up
Sat Nov 14 21:39:55 CET 2015	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89	Top

- Raw data is not only numbers, data can contain strings and timestamps.
- Can be entirely categorical or numerical, or a mixture of both.
- We do not always get tidy data (clean data). Cleaning and manipulating the data can be a challenge.

Raw data

Sat Nov 14 21:39:05 CET 2015	10.0	8.04	10.0	9.14	10.0	7.46	8.0	**6.58**	Left
Sat Nov 14 21:39:10 CET 2015	8.0	**6.95**	8.0	8.14	8.0	**6.77**	8.0	5.76	Left
Sat Nov 14 21:39:15 CET 2015	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71	Right
Sat Nov 14 21:39:20 CET 2015	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84	Top
Sat Nov 14 21:39:25 CET 2015	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47	Top
Sat Nov 14 21:39:30 CET 2015	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04	Bottom
Sat Nov 14 21:39:35 CET 2015	**6.0**	7.24	**6.0**	**6.13**	**6.0**	**6.08**	8.0	5.25	Bottom
Sat Nov 14 21:39:40 CET 2015	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50	Right
Sat Nov 14 21:39:45 CET 2015	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56	Up
Sat Nov 14 21:39:50 CET 2015	7.0	4.82	7.0	7.26	7.0	**6.42**	8.0	7.91	Up
Sat Nov 14 21:39:55 CET 2015	5.0	5.68	5.0	4.74	5.0	5.73	8.0	**6.89**	Top

- By simply using a visual encoder new information pops out.
- By just looking at the new information we can infer something about the data.
- So far we have not graphed the data.

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

Do you spot any correlation or peculiarity on these datasets?

Anscombe's quartet

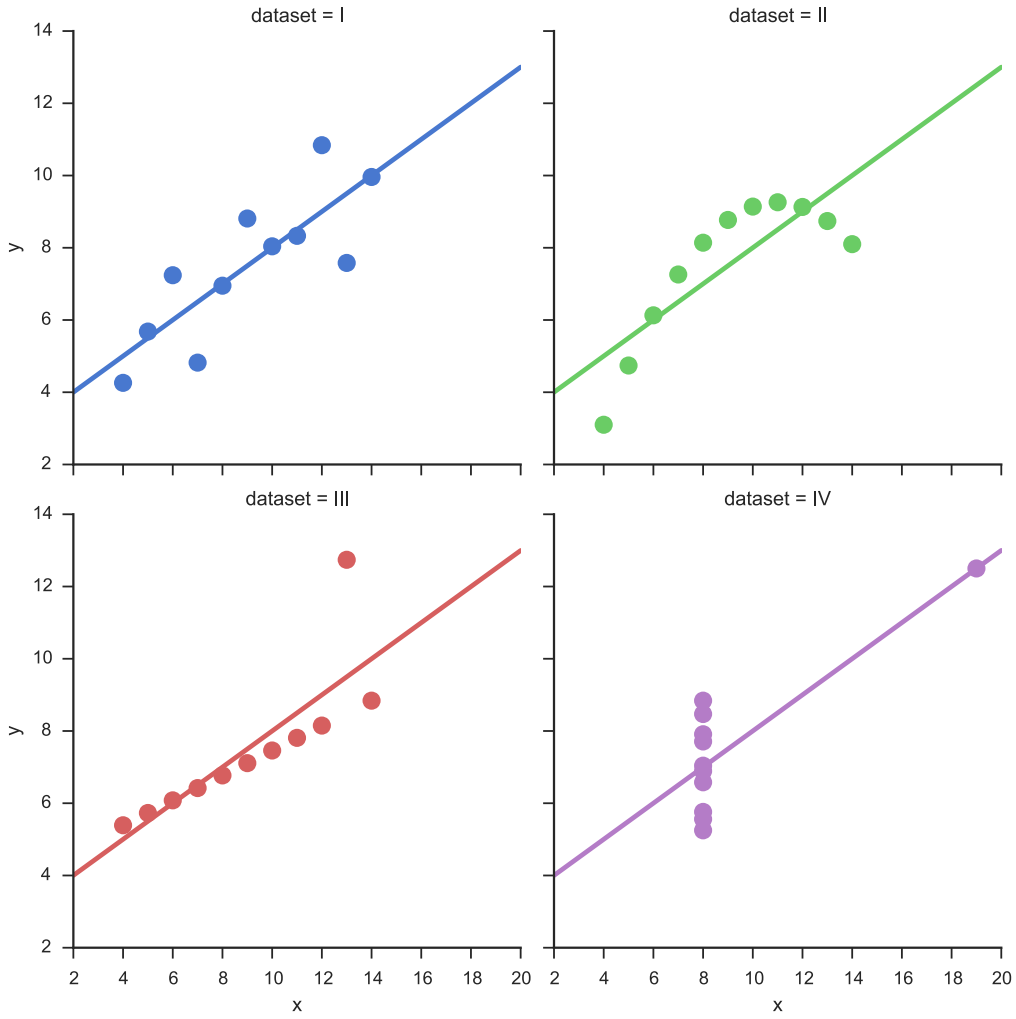
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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

For all datasets:

Statistical property	Value
Sample size	11
Mean (x)	9
Variance (x)	11
Mean (y)	7.50
Variance (y)	4.122
Correlation	0.816
Linear regression	Y = 3.00 + 0.5000X

Even if the four datasets are different, they have nearly identical simple statistical properties. What will we see when the data is graphed?

Anscombe’s quartet



For all datasets:

Statistical property	Value
Sample size	11
Mean (x)	9
Variance (x)	11
Mean (y)	7.50
Variance (y)	4.122
Correlation	0.816
Linear regression	Y = 3.00 + 0.5000X

Anscombe's quartet comprises four datasets that have nearly identical simple statistical properties, yet appear very different when graphed.

The how of data visualization and visual storytelling

Data visualization tools

- Factors to consider when choosing a data visualization tool:
  - Ease of use.
  - Flexibility (modifiable, configurable and extensible).
  - Reusability.
  - Interactivity.
  - Expressiveness.
  - Aesthetics.
  - Portability.
  - Accessibility (price tag).

Data visualization tools

- A few options available (commercial and open-source):
  - IBM Watson analytics, Microsoft azure, Amazon web services analytics, Oracle big data discovery, Google Cloud Platform.
  - Mathematica, matlab, scilab, octave.
  - Minitab, SAS, qlik, tableau, gapminder, polestar.
  - Excel, libreoffice.
  - OpenGL, VTK, Java2D, processing.
  - Javascript and JS charting libraries (Google charts, plotly, D3.JS, highcharts JS, chartJS, ember charts).
  - Python, R.

Data visualization tools

Our approach

- A web-based interactive data visualization and analysis toolkit that runs:
    - On the client side: javascript, D3.JS, webGL and html5.
    - On the server side: node.js, Python, R and shiny.

- We speak the language of the web.
- We are able to control every pixel of the screen.
- The server tools give us access to extensive and advanced data analytics capabilities.

Web-based interactive data visualization and analysis toolkit

- We aim at enabling users to manipulate, analyze and visualize their data interactively.
- As the tools are implemented using the language of the web (javascript and html5), they can run from any device with a working web browser (PC, tablet, smart-phone).
- The learning curve is minimal as the user only needs to interact with the web browser interface.
- The tool supports DSV, JSON, XML and SQL format.

`https://github.com/joelguerrero/dae4cfd`

Toolkit for Data Analytics
& Design-Analysis of Experiments

DA^{2}e4 CFD

Web-based interactive data visualization and analysis toolkit

- The data can be obtained from any discipline (social sciences, econometrics, marketing, the social web, sports, health care, bioinformatics, engineering, etc.) or the user's daily activity (blood pressure measurement, time to arrive to your workplace, daily calories intake, etc).
- But our main goal is to work with data obtained from numerical simulations and optimization studies.
- **Visual storytelling for CFD.**

`https://github.com/joelguerrero/dae4cfd`

Toolkit for Data Analytics
& Design-Analysis of Experiments

DA^{2}e4 CFD

Web-based interactive data visualization and analysis toolkit

- **Data visualization and exploration.**
  - Plotting of multidimensional data.
  - Machine learning and predictive analytics.
  - Identifying outliers.
  - Cross-filtering data.
  - Summary statistics (but do not just rely on this).

- **Interactive visualization.**
  - Facilitates manipulation and exploration of the data.
  - Allows cross-filtering of data.
  - Interactive tools are great for collaboration between groups.
  - Enable summaries with access to the details.

- **Reports and data communication.**
  - Living documents.
  - It is a great tool for show and tell a compelling story to all audiences.

- **More exploration, more connections, more insight.**

“In God we trust, all others bring data.”

— William E. Deming

“Show data variation and not design variation.”

— Edward Tufte

Dataset 1

Design space exploration and optimization dataset

This dataset and the associated scripts (html and js) are available at the following github link:

```
https://github.com/joelguerrero/joelguerrero.github.io/tree/master/p1/
```

The main github.io page is:

```
http://joelguerrero.github.io/p1/
```

To run the web-based version go to the following address:

```
http://joelguerrero.github.io/p1/dataset1/1_boxplot
http://joelguerrero.github.io/p1/dataset1/2_histograms
http://joelguerrero.github.io/p1/dataset1/3_regression
http://joelguerrero.github.io/p1/dataset1/4_hexbin
```

Dataset 1

Column 1	Column 2	Column 3	Column 4	Column 5
3.95E+00	1.63E+00	0.44242	0.0055127	0.0666262
7.11E+00	3.44E+00	0.860229	0.00615626	0.174343
6.90E+00	1.78E+00	0.782869	0.0089705	0.122455
7.79E+00	3.82E+00	0.960738	0.00650422	0.203037
1.25E+00	3.28E+00	0.158067	0.00373114	0.031634
5.63E+00	2.11E+00	0.643855	0.00648422	0.107407
…	…	…	…	…
N	N	N	N	N

- Data obtained from a design space exploration study.
- Sample size: 5 X 121
- All the data is numerical.
- Tidy data.

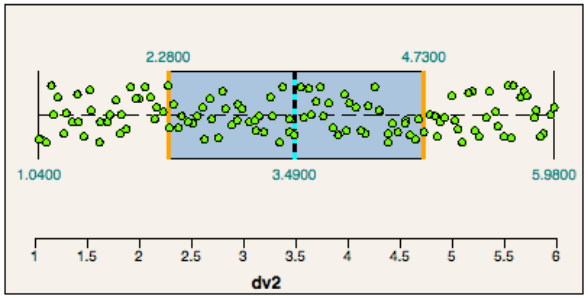
How do I explore this dataset or any dataset?

How do I explore this dataset or any dataset?

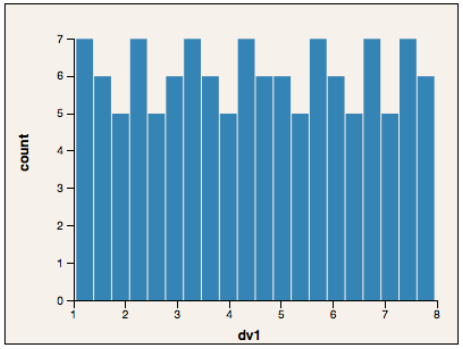
- The first step is getting the data and cleaning the data.
- If I have tidy data I am done.
- If I need to do data wrangling, my time is divided a follows:
    - 50% preparing the data.
    - 40% complaining about the data.
    - 9% for actual programming
    - 1% thinking how to best visualize the data.

How do I explore this dataset or any dataset?

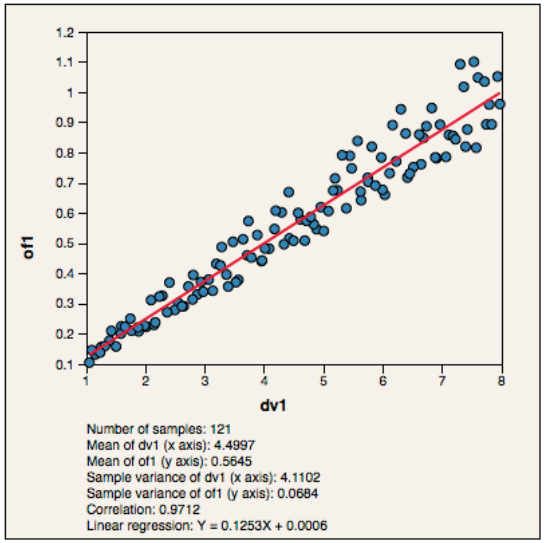
Boxplot



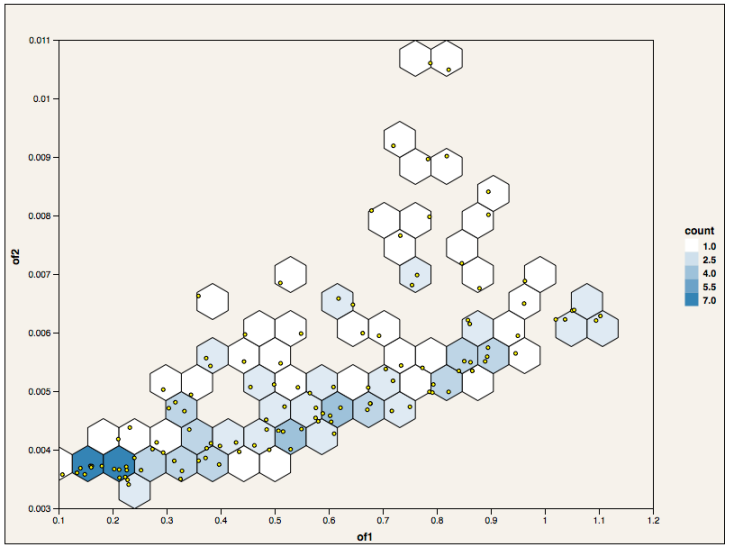
Histogram



Scatter plot + regression

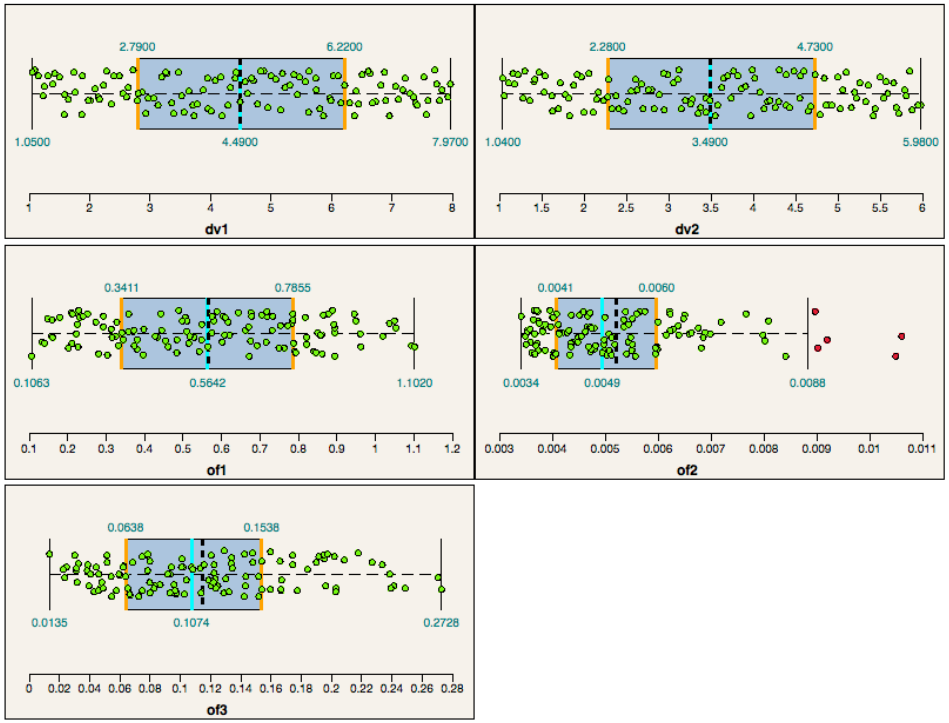


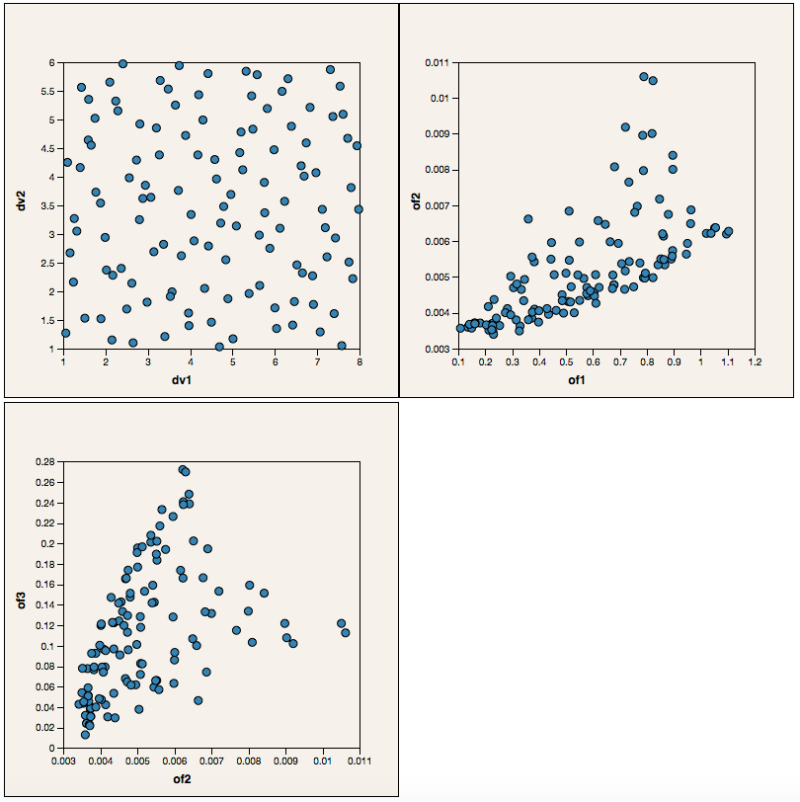
Hexbin



How do I explore this dataset or any dataset?

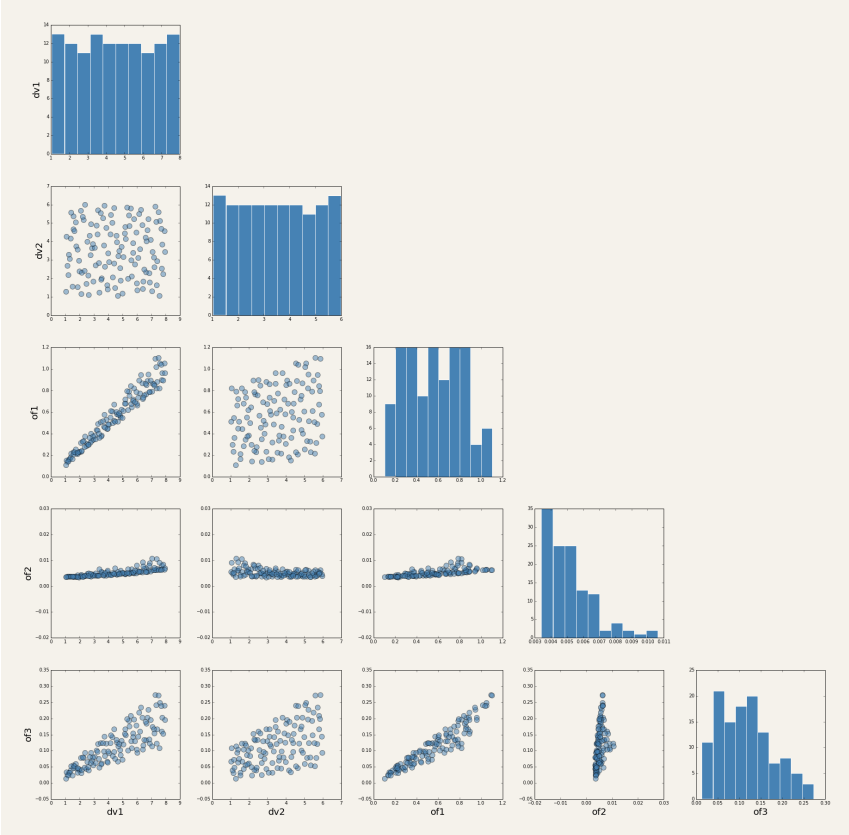
Small multiples (univariate data) and paired-data (bivariate data) plots

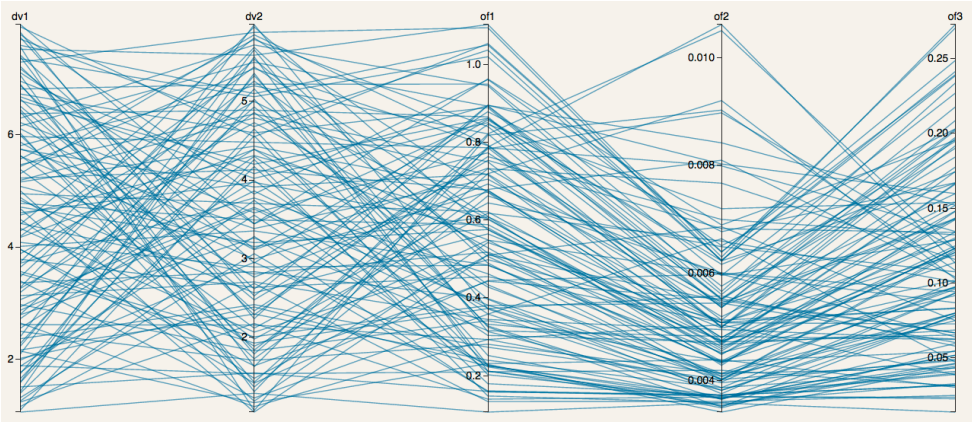




How do I explore this dataset or any dataset?

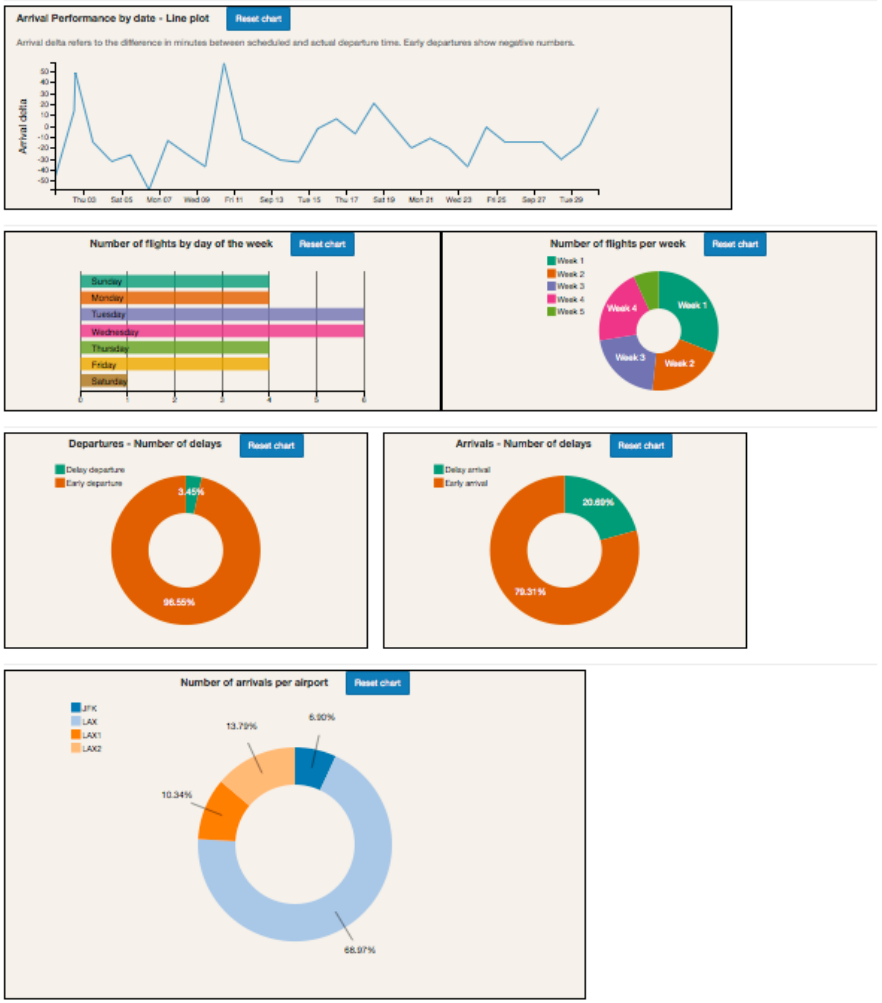
Scatter matrix plot and parallel coordinates for multivariate data





How do I explore this dataset or any dataset?

Interactive dashboard for cross filtering data and exploring multivariate data



Dataset 2

Design space exploration and optimization dataset

This dataset and the associated scripts (html and js) are available at the following github link:

```
https://github.com/joelguerrero/joelguerrero.github.io/tree/master/p1/
```

The main github.io page is:

```
http://joelguerrero.github.io/p1/
```

To run the web-based version go to the following address:

```
http://joelguerrero.github.io/p1/dataset2/1_boxplot
http://joelguerrero.github.io/p1/dataset2/2_histograms
http://joelguerrero.github.io/p1/dataset2/3_regression
http://joelguerrero.github.io/p1/dataset2/4_hexbin
http://joelguerrero.github.io/p1/dataset2/5_parallel_coordinates
http://joelguerrero.github.io/p1/dataset2/6_pareto_scatter
```

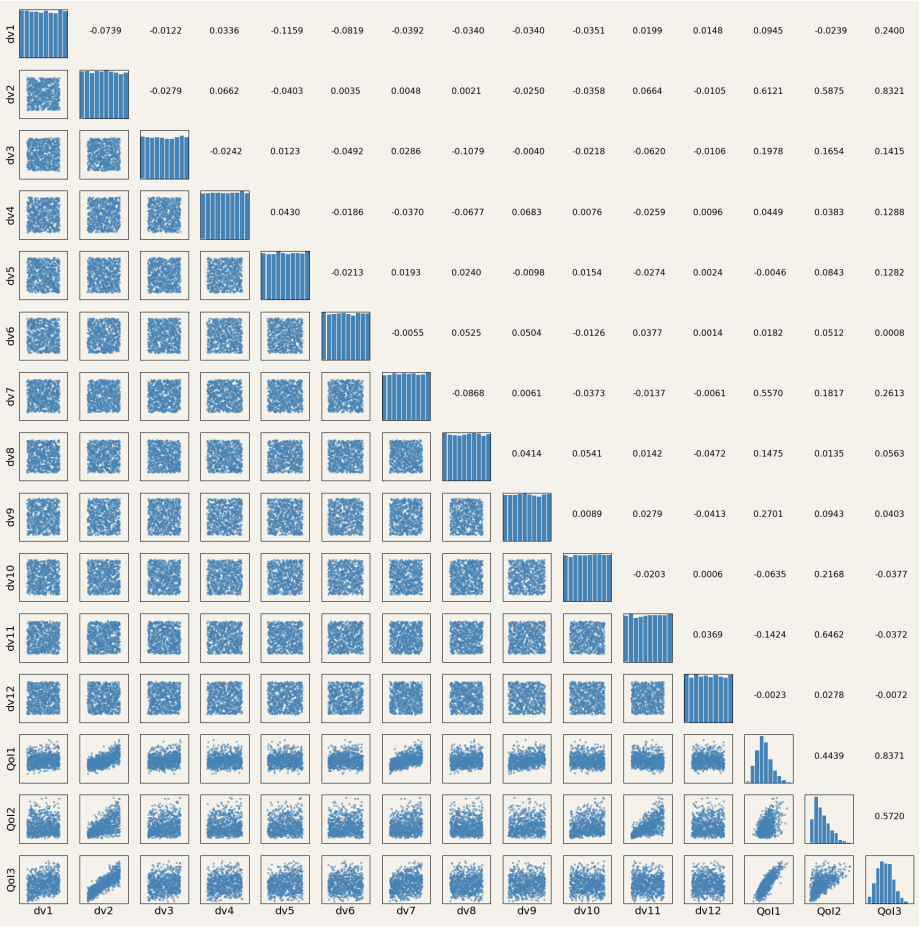
Dataset 2

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
4.87E-01	1.43E-01	5.90E-01	7.99E-01	8.64E-01	6.29E-01	2.08E-01	2.28E-01	2.58E-01	7.49E+00	3.88E+01	-1.34E+01	35.2251	122.277	339.623
2.85E-01	1.43E-01	6.49E-01	6.33E-02	8.88E-01	9.87E-01	3.09E-01	2.24E-01	1.66E-01	1.32E+00	9.31E-01	-4.96E-01	42.1183	4.13759	309.449
3.49E-01	5.40E-01	1.93E-01	4.72E-01	6.07E-02	3.80E-01	2.60E-01	2.02E-01	1.47E-01	6.05E+00	2.95E+00	-4.68E+00	41.7185	37.3891	467.299
3.07E-01	4.09E-01	5.77E-01	1.81E-02	1.97E-01	2.30E-01	3.44E-01	2.36E-01	2.02E-01	3.21E+00	3.34E+01	-1.17E+01	45.6606	155.98	484.209
6.53E-01	1.56E-02	5.53E-01	9.93E-02	6.88E-01	2.09E-01	3.05E-01	3.46E-01	1.50E-01	8.64E+00	1.12E-01	-6.07E+00	45.3468	30.2206	291.584
3.09E-01	7.93E-01	3.16E-01	4.10E-01	9.95E-01	2.74E-01	3.26E-01	2.82E-01	1.88E-01	3.95E+00	6.59E+00	-2.71E+00	55.0919	79.864	930.964
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

- Data obtained from a design space exploration study.
- Sample size: 15 X 777
- All the data is numerical.
- Not so tidy data.

Dataset 2

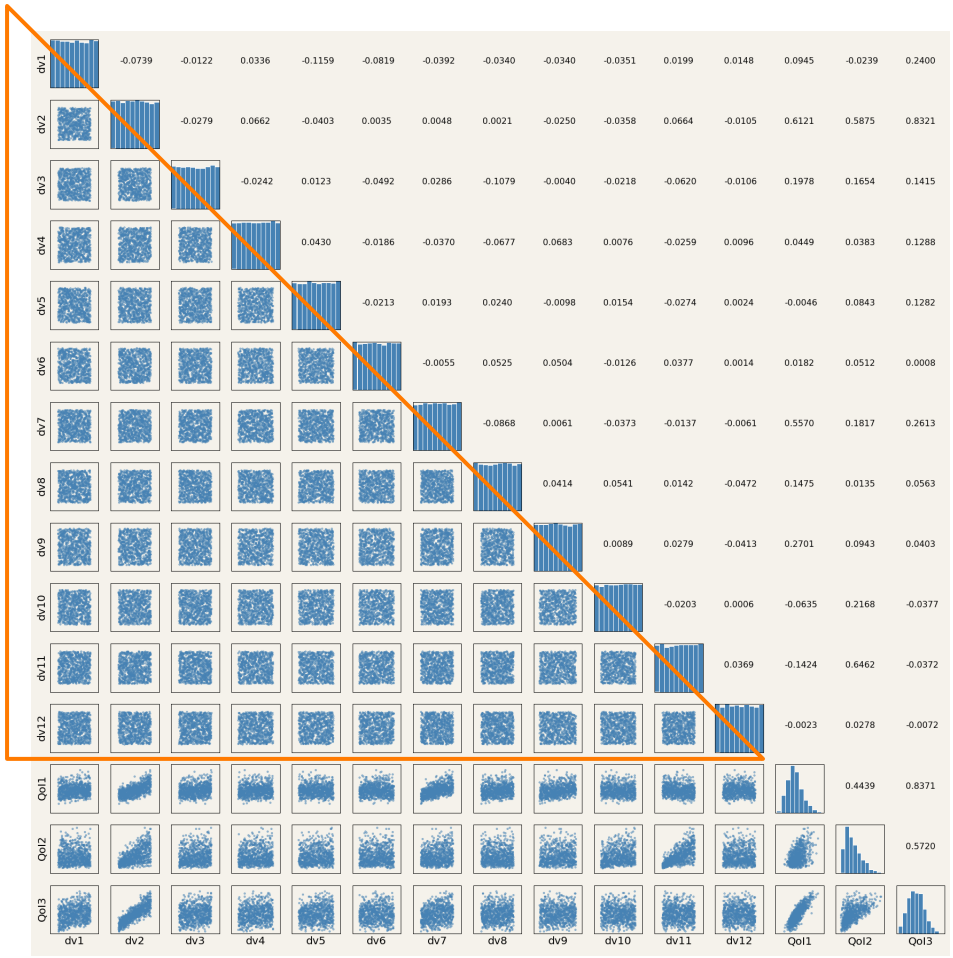
Scatter matrix plot of a design space exploration study
This plot shows correlation, skewness, kurtosis, tendency and distribution of the data



Dataset 2

Scatter matrix plot of a design space exploration study

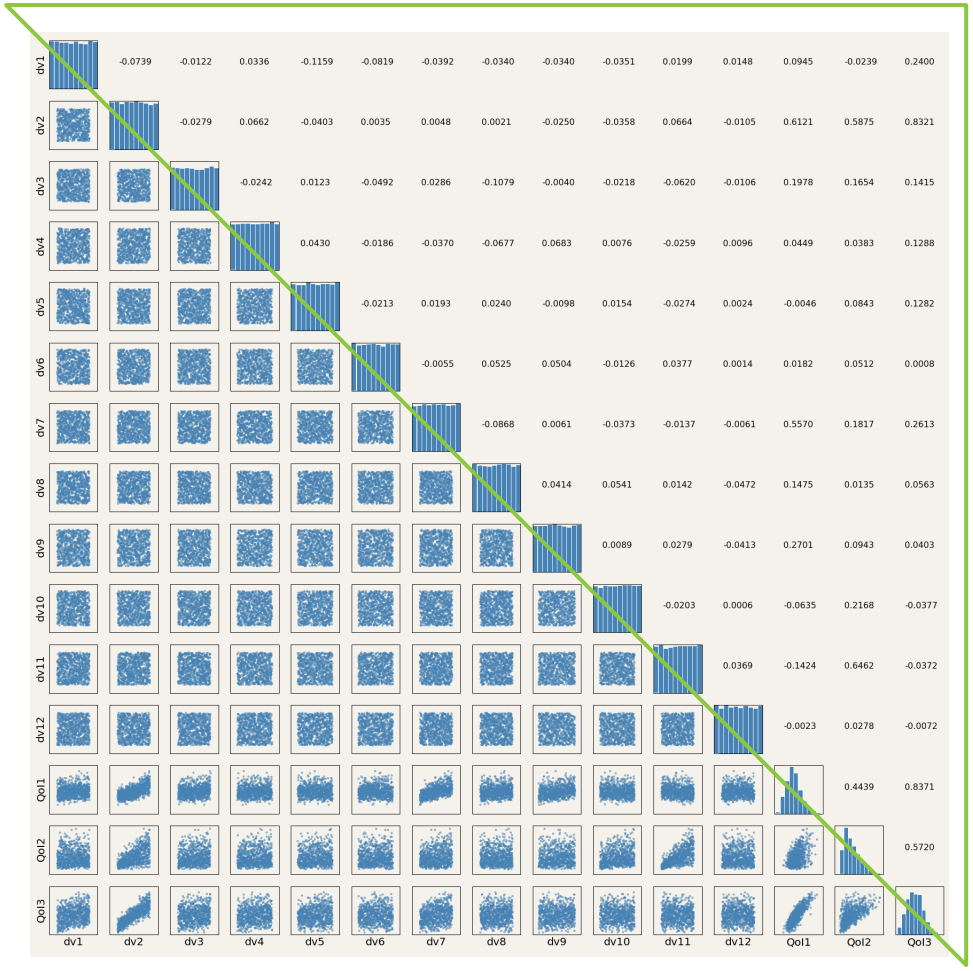
Scatter plot of design variables distribution (sampling distribution in design space)



Dataset 2

Scatter matrix plot of a design space exploration study

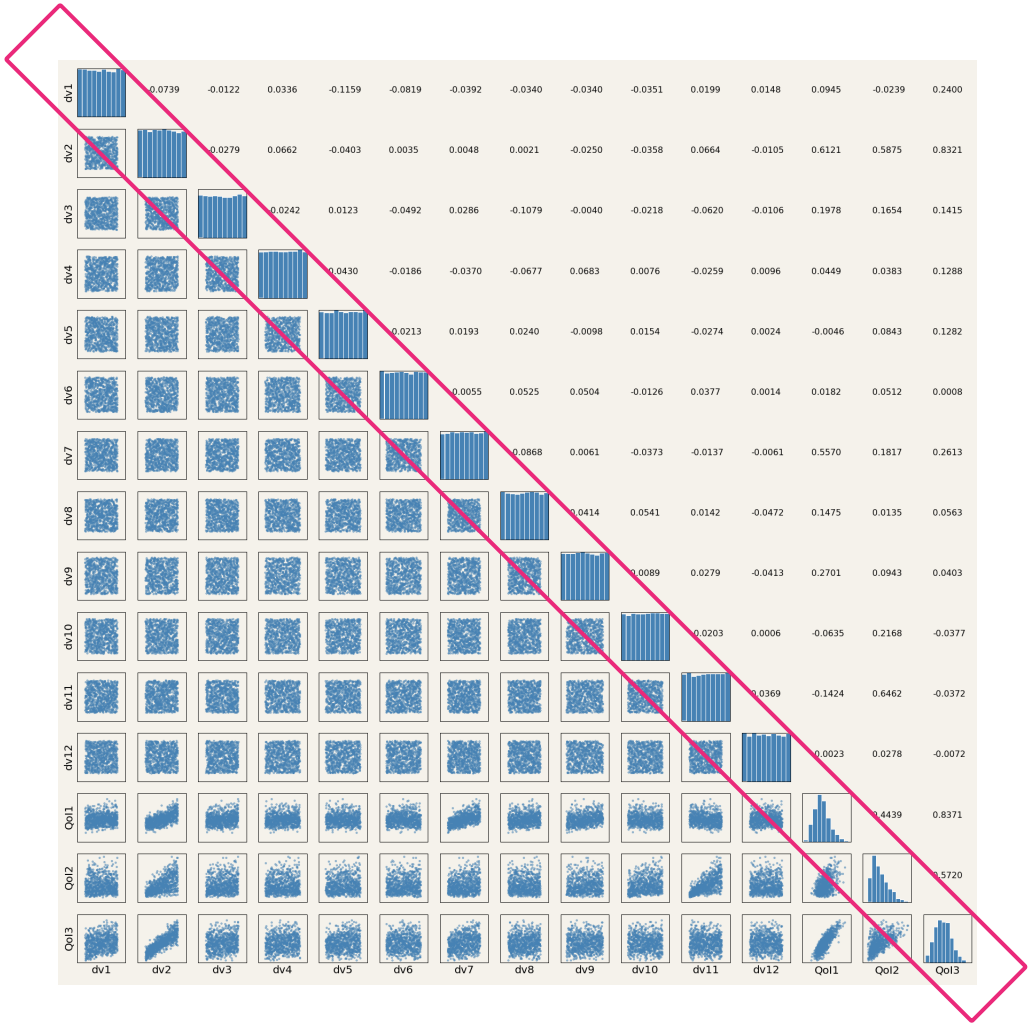
Correlation matrix of design space variables (design variables and objective functions)



Dataset 2

Scatter matrix plot of a design space exploration study

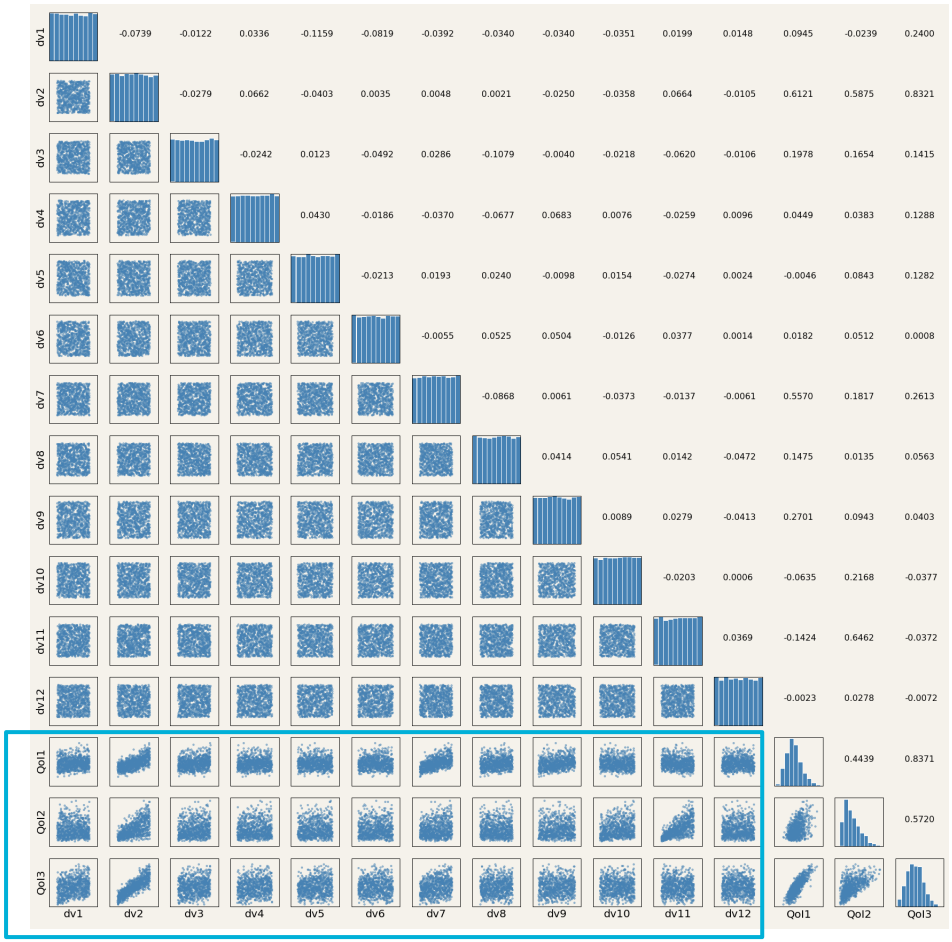
Histograms of design variables and objective functions



Dataset 2

Scatter matrix plot of a design space exploration study

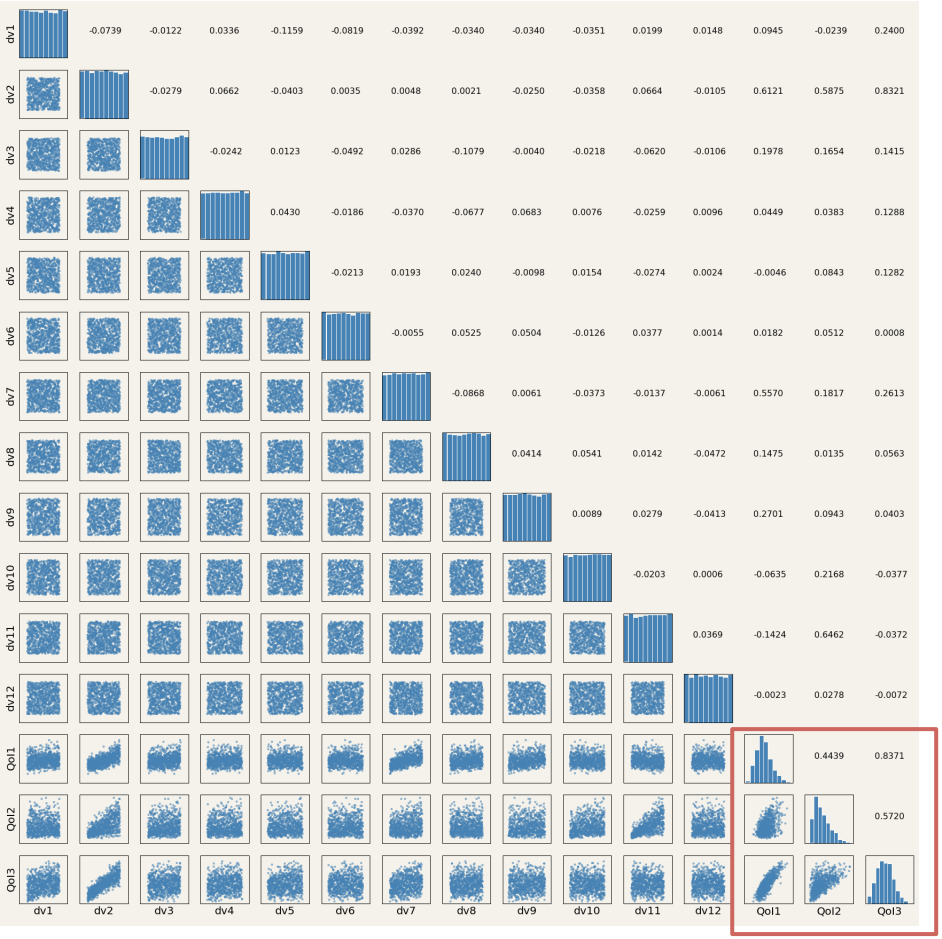
Response of design space (scatter plot of design variables vs. objective functions)



Dataset 2

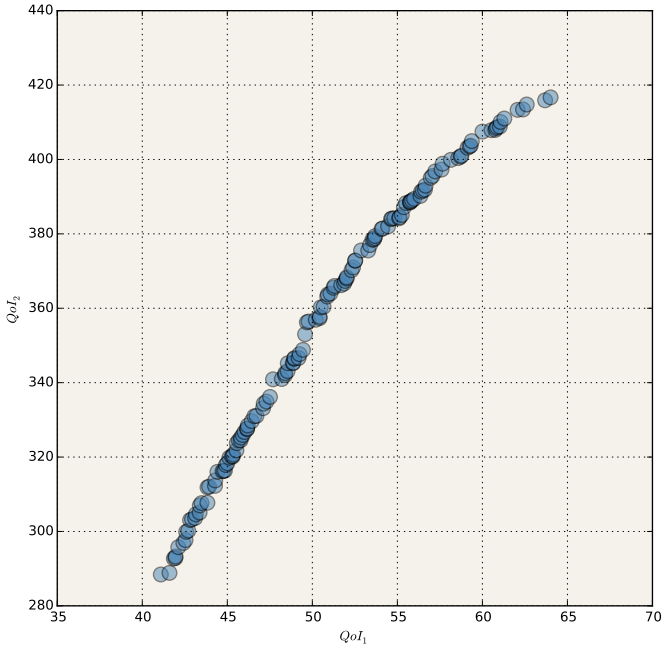
Scatter matrix plot of a design space exploration study

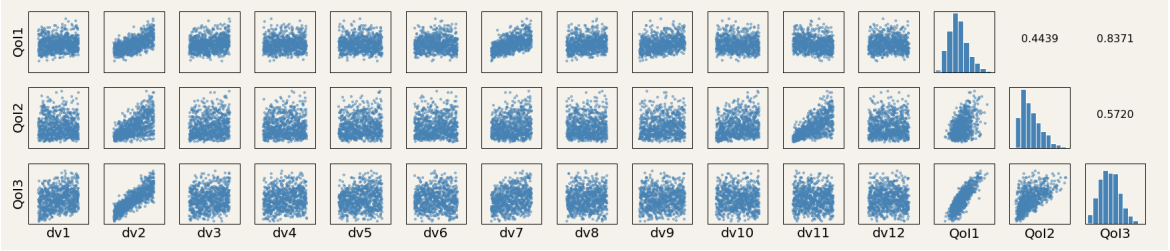
Response or trade-off of objective functions



Dataset 2

Pareto front and overall response of the design space.

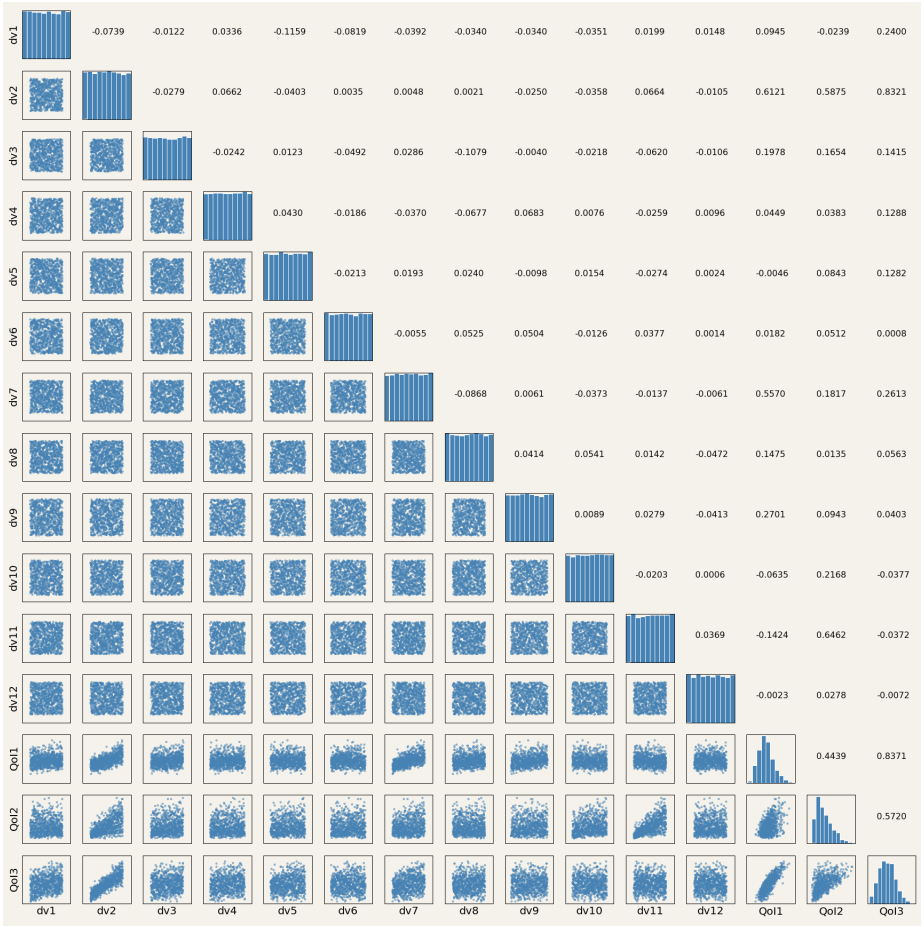




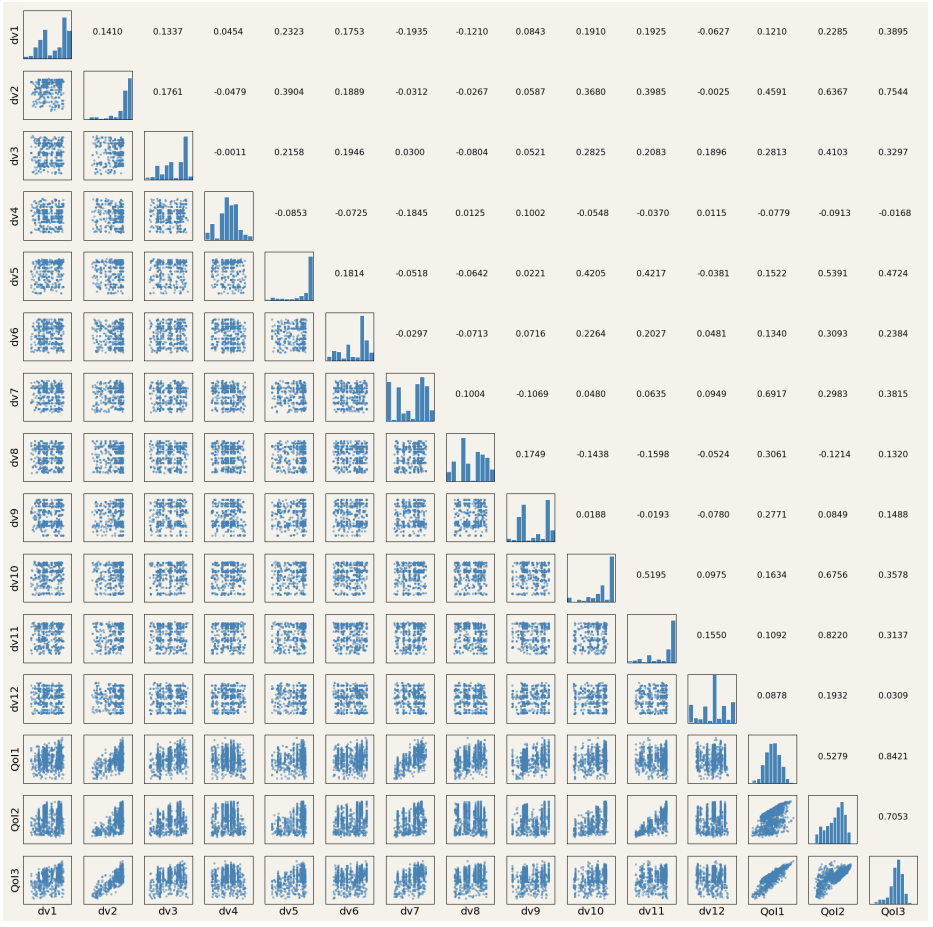
- The Pareto front was constructed using surrogate based optimization.
- How do we relate the trade-offs in the Pareto front with the design variables?

Dataset 2

Comparison of two scatter matrix plots



Design space exploration study

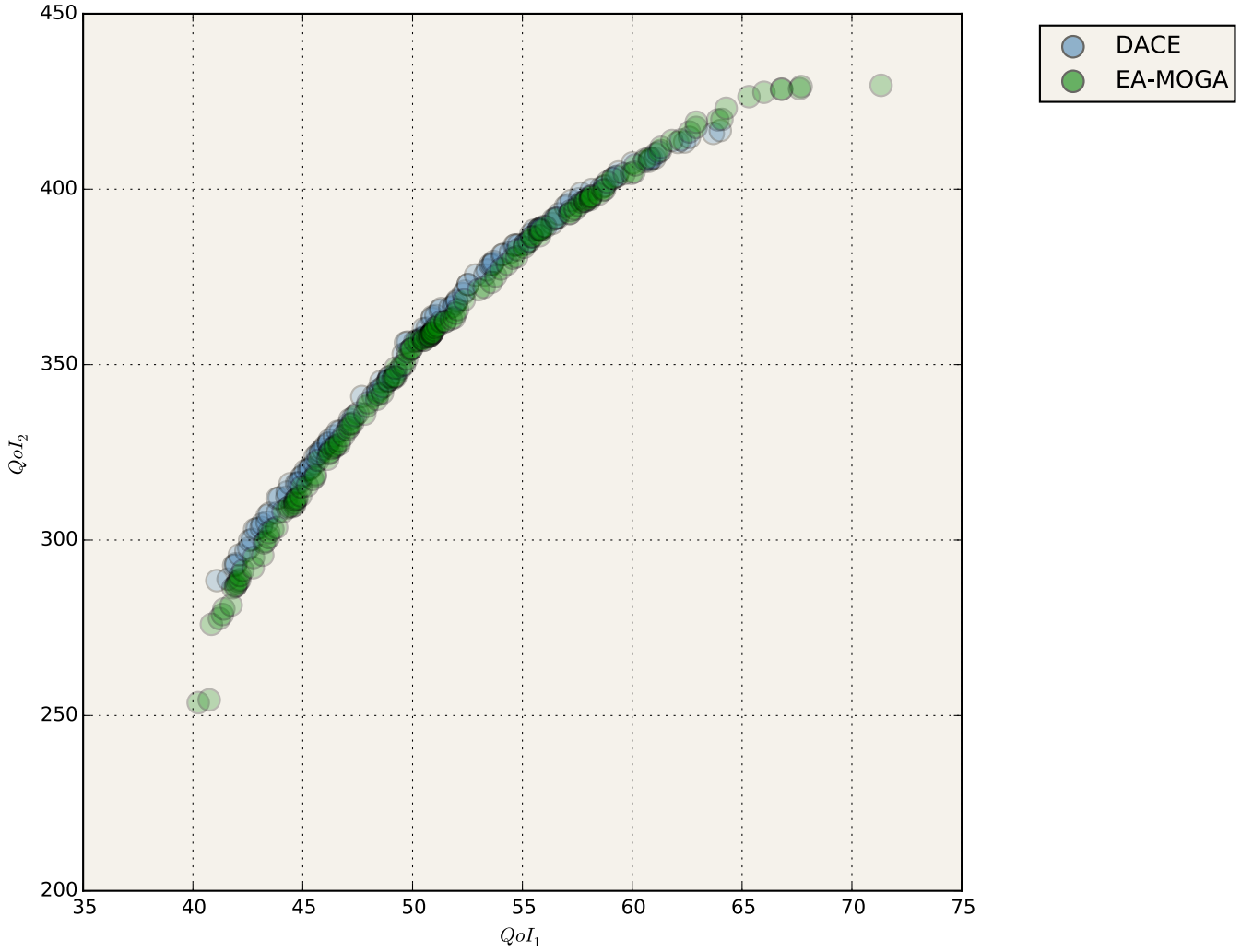


Multi-objective optimization study

- These two experiments give very similar Pareto fronts, however, when we conduct EDA we clearly see that the sampling, correlations and trends are different.

Dataset 2

Pareto front comparison



Dataset 3

Effectiveness of three antibiotics against 16 different bacteria dataset

This dataset and the associated scripts (html and js) are available at the following github link:

```
https://github.com/joelguerrero/joelguerrero.github.io/tree/master/p1/
```

The main github.io page is:

```
http://joelguerrero.github.io/p1/
```

To run the web-based version go to the following address:

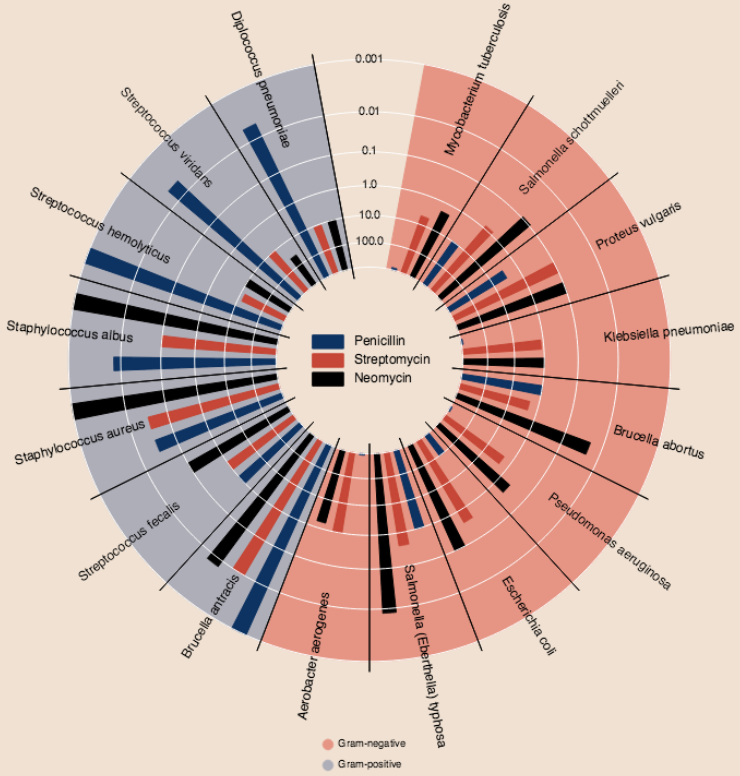
```
http://joelguerrero.github.io/p1/dataset3/1_boxplot
http://joelguerrero.github.io/p1/dataset3/2_scatterplot
```

Dataset 3

Bacteria	Penicillin	Streptomycin	Neomycin	Gram stain
Aerobacter aerogenes	870	1	1.6	neg
Brucella abortus	1	2	0.02	neg
Brucella antracis	0.001	0.01	0.007	pos
Diplococcus pneumoniae	0.005	11	10	pos
Escherichia coli	100	0.4	0.1	neg
Klebsiella pneumoniae	850	1.2	1	neg
Mycobacterium tuberculosis	800	5	2	neg
Proteus vulgaris	3	0.1	0.1	neg
Pseudomonas aeruginosa	850	2	0.4	neg
Salmonella (Eberthella) typhosa	1	0.4	0.008	neg
Salmonella schottmuelleri	10	0.8	0.09	neg
Staphylococcus albus	0.007	0.1	0.001	pos
Staphylococcus aureus	0.03	0.03	0.001	pos
Streptococcus fecalis	1	1	0.1	pos
Streptococcus hemolyticus	0.001	14	10	pos
Streptococcus viridans	0.005	10	40	pos

- Studies on antibiotic synergism and antagonism. J. Bacteriol. 1952 Jul; 64(1): 29–39
- Sample size: 5 X 16
- Numerical data and categorical.
- Tidy data.

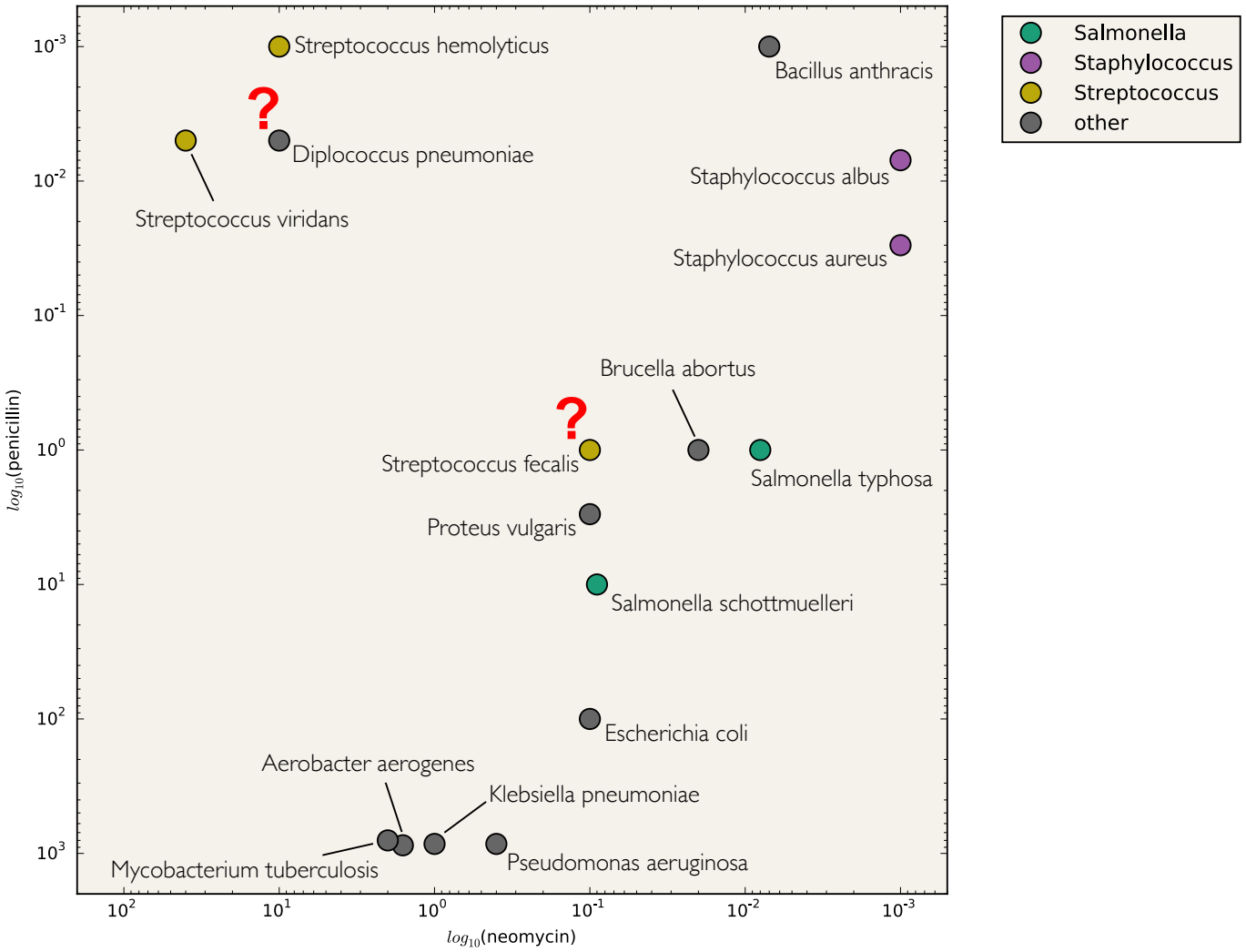
Dataset 3



Bacteria	Penicillin	Streptomycin	Neomycin	Gram stain
Aerobacter aerogenes	870	1	1.6	-
Brucella abortus	1	2	0.02	-
Brucella antracis	0.001	0.01	0.007	+
Diplococcus pneumoniae	0.005	11	10	+
Escherichia coli	100	0.4	0.1	-
Klebsiella pneumoniae	850	1.2	1	-
Mycobacterium tuberculosis	800	5	2	-
Proteus vulgaris	3	0.1	0.1	-
Pseudomonas aeruginosa	850	2	0.4	-
Salmonella (Eberthella) typhosa	1	0.4	0.008	-
Salmonella schottmuelleri	10	0.8	0.09	-
Staphylococcus albus	0.007	0.1	0.001	+
Staphylococcus aureus	0.03	0.03	0.001	+
Streptococcus fecalis	1	1	0.1	+
Streptococcus hemolyticus	0.001	14	10	+
Streptococcus viridans	0.005	10	40	+

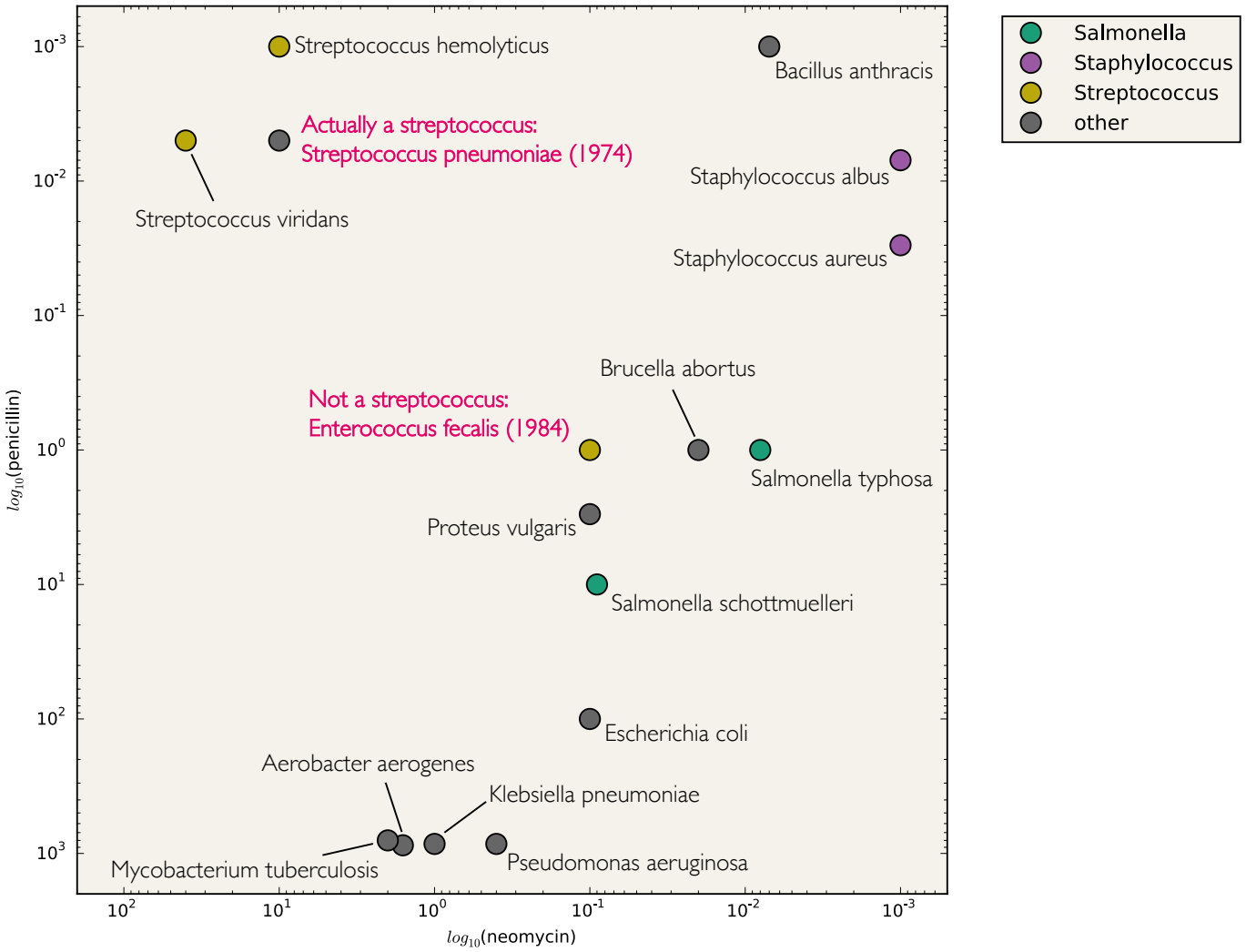
- This celebrated plot created by Will Burtin in 1951 illustrates the effectiveness of three antibiotics against 16 different bacteria. The bars represent minimum inhibitory concentrations (MICs). The longer the bar, the smaller the effective dose.
- This figure address the following question, how the drugs compare against each other for each bacteria?
- However, by looking at the figure is not possible to answer the following question, how do the bacteria group together in response to the drugs?

Dataset 3



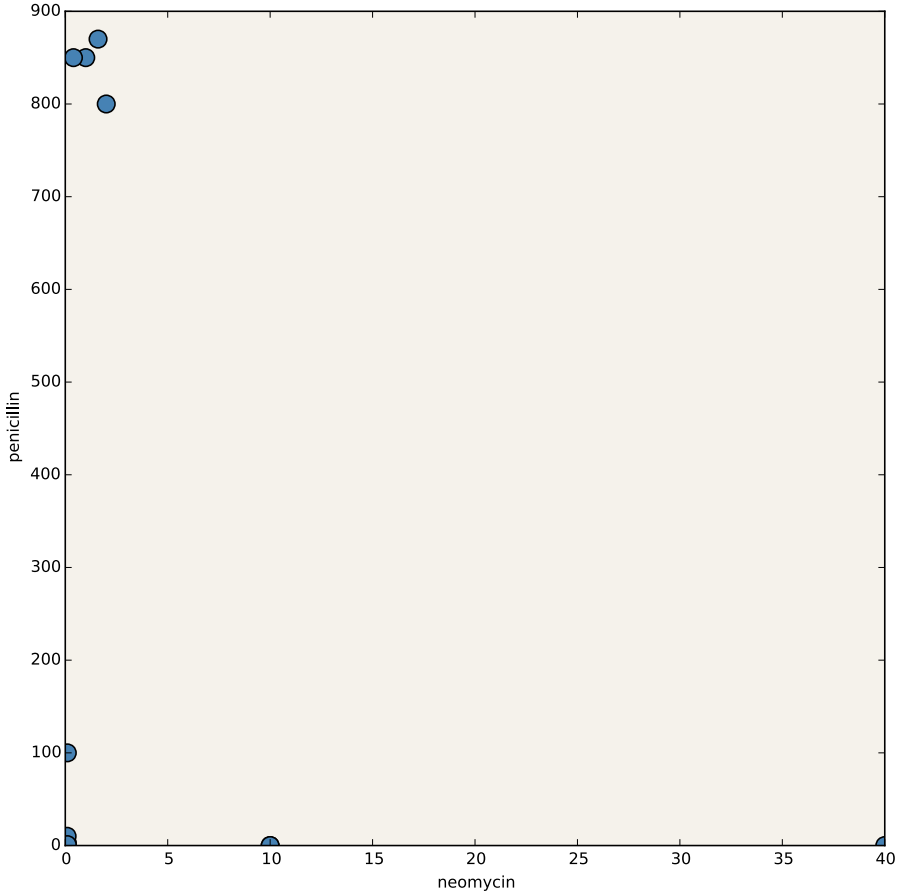
How do the bacteria group together in response to penicillin and neomycin?

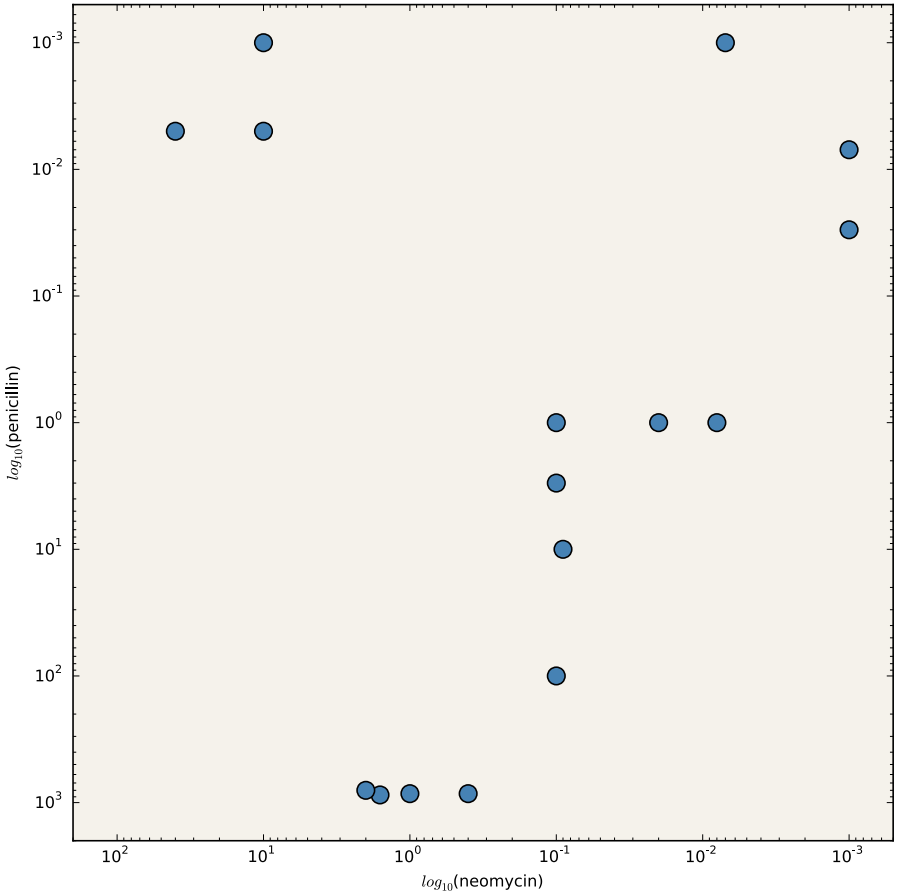
Dataset 3



How do the bacteria group together in response to penicillin and neomycin?

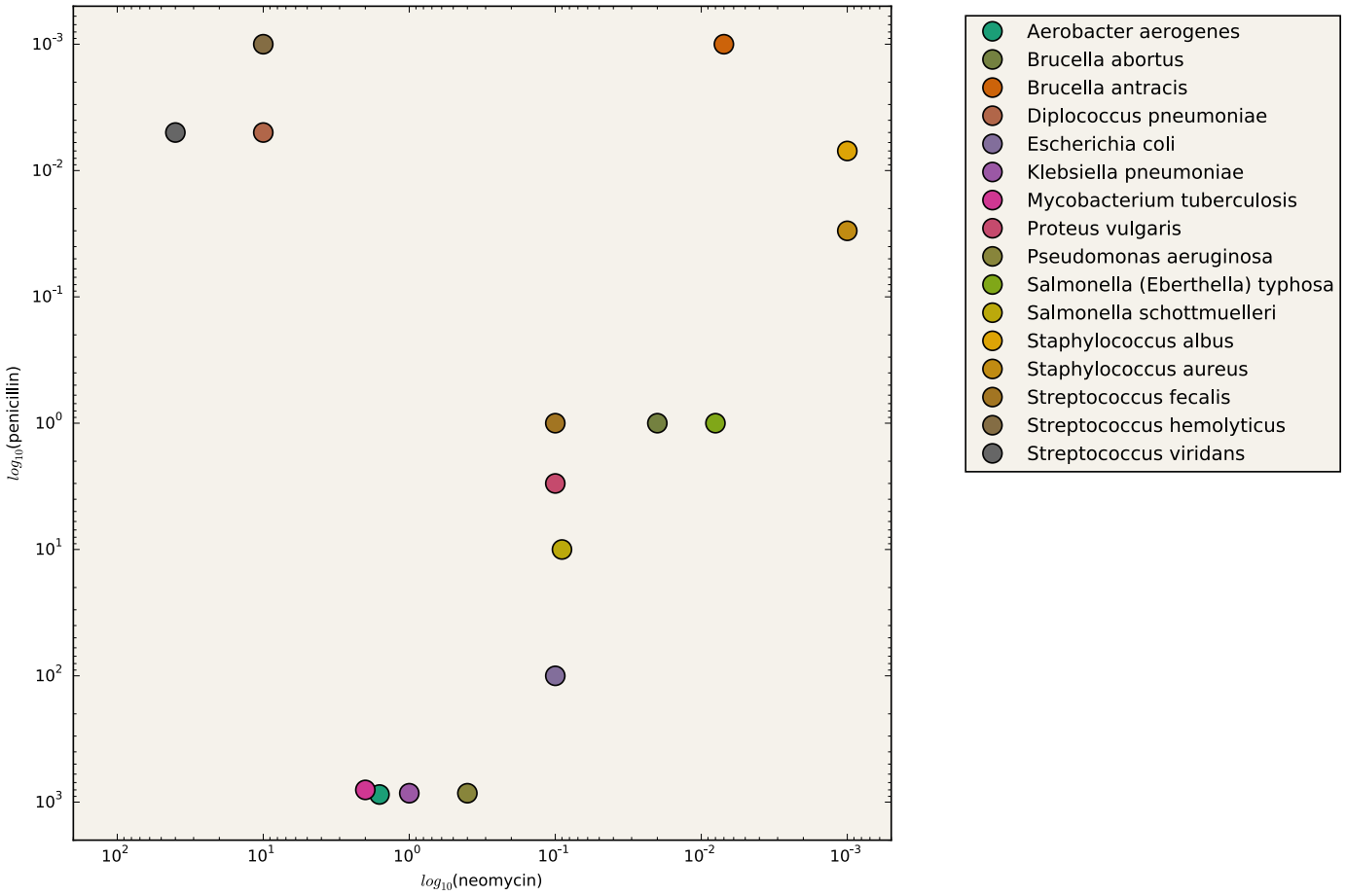
Dataset 3





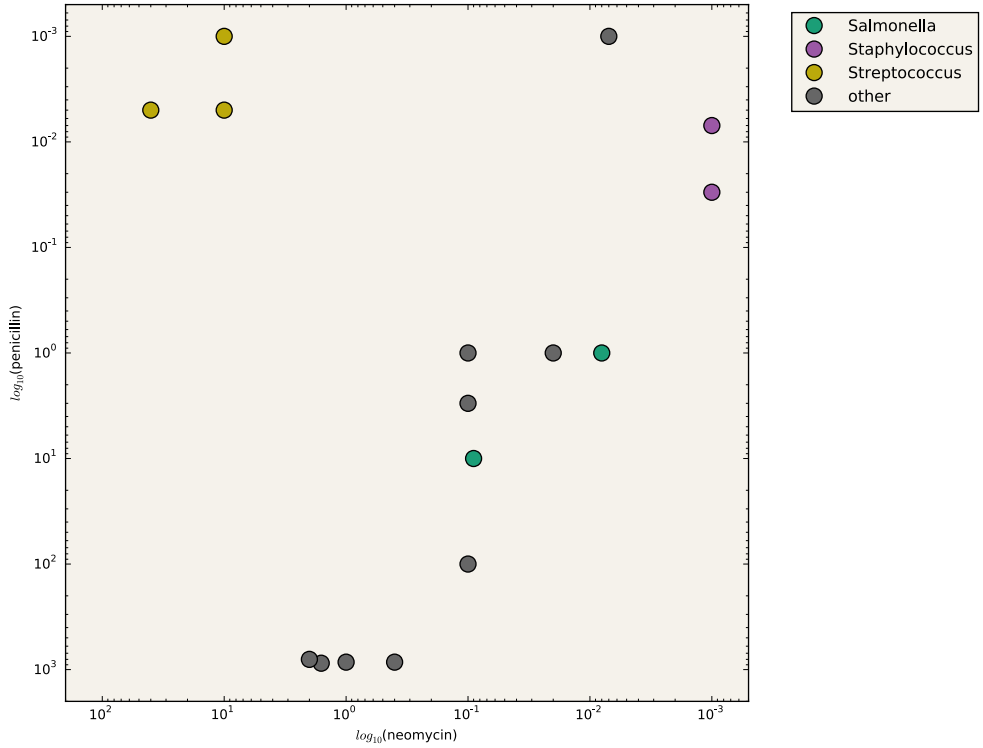
Scales and domain dimensions are important.

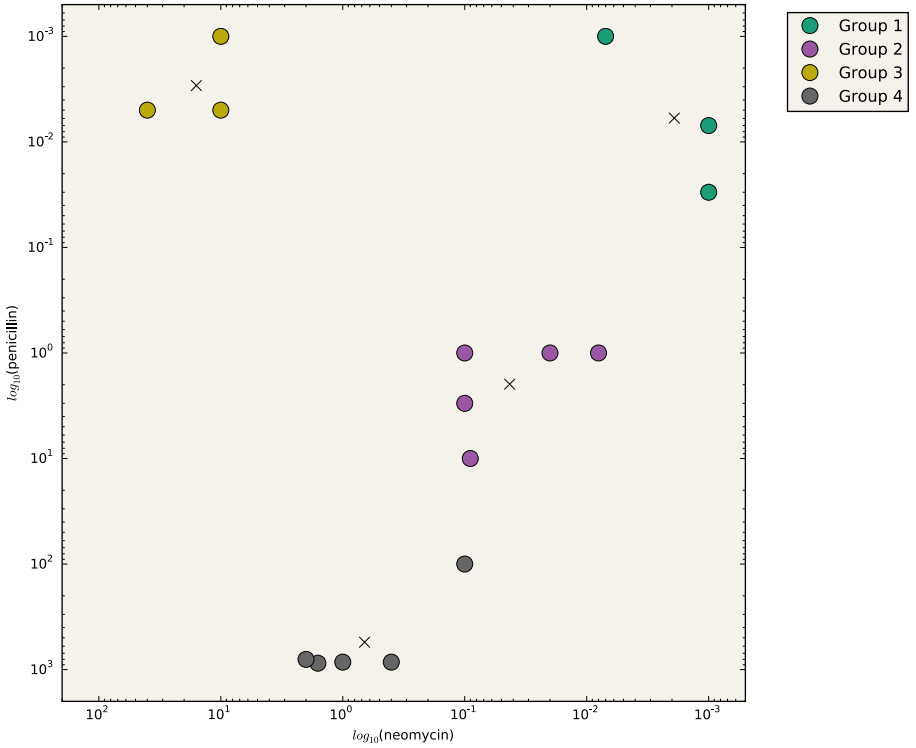
Dataset 3



As well as visual encoders and using legends.

Dataset 3





- By using a clustering method (k means clustering), we can identify how the bacteria groups.
- The information is in the data.

Dataset 4

Sport analytics dataset

This dataset and the associated scripts (html and js) are available at the following github link:

```
https://github.com/joelguerrero/joelguerrero.github.io/tree/master/p1/
```

The main github.io page is:

```
http://joelguerrero.github.io/p1/
```

To run the web-based version go to the following address:

```
http://joelguerrero.github.io/p1/dataset4/1_scatterplot
http://joelguerrero.github.io/p1/dataset4/2_boxplot
```

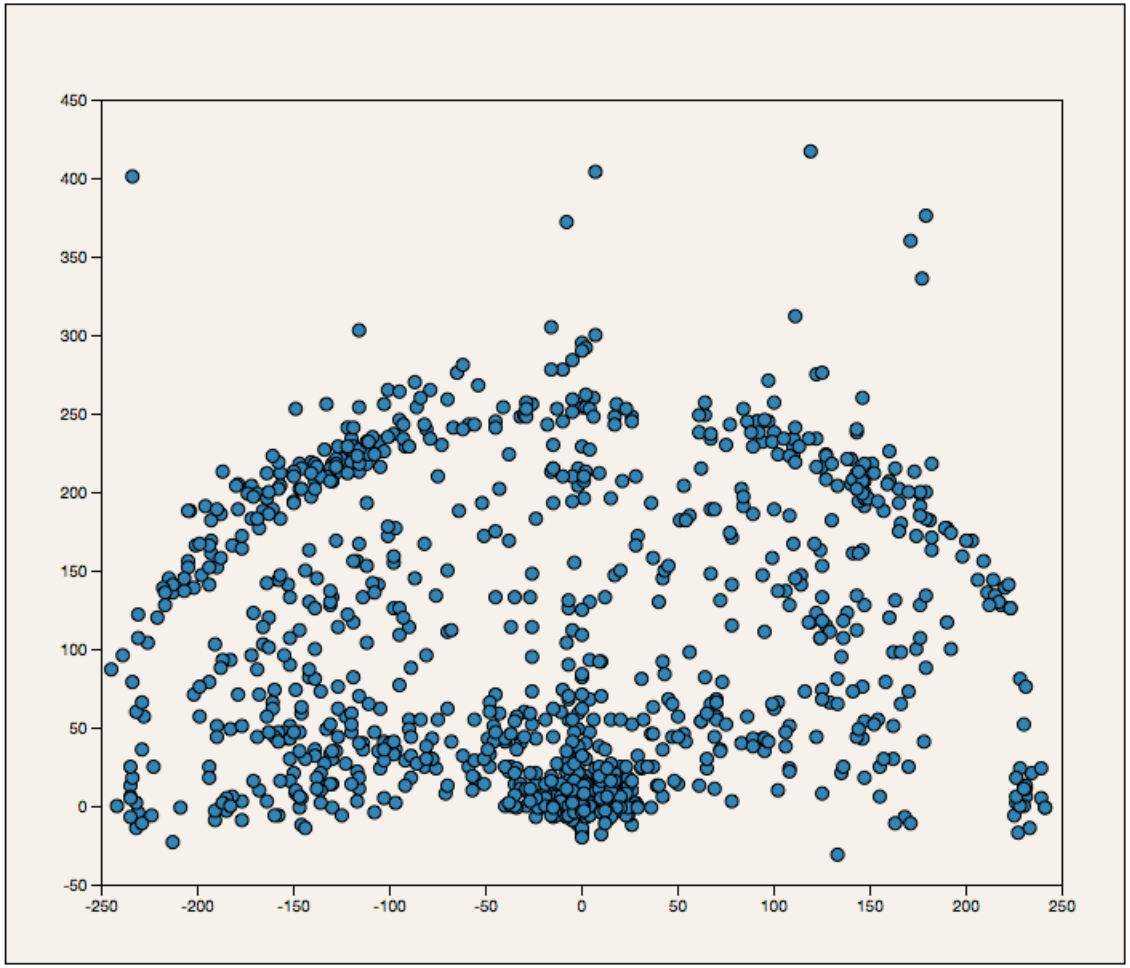
Dataset 4

C1	C2	C3	C4-C7	C8	C9	C10	C11-C16	C17	C18	C19	C20	C21
Categorical data	21400018	4	Categorical data	1	11	20	Categorical data	18	114	148	1	0
Categorical data	21400018	33	Categorical data	1	6	30	Categorical data	0	-7	0	1	1
Categorical data	21400018	53	Categorical data	1	4	45	Categorical data	12	-105	63	1	0
Categorical data	21400018	77	Categorical data	1	2	31	Categorical data	22	227	-16	1	0
Categorical data	21400018	82	Categorical data	1	1	51	Categorical data	26	91	246	1	0
Categorical data	21400018	136	Categorical data	2	9	18	Categorical data	9	70	58	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...
N	N	N	N	N	N	N	N	N	N	N	N	N

- Data scrapped from the web.
- Sample size: 21 X 1279
- Numerical data, categorical data and timestamps.
- Tidy data.

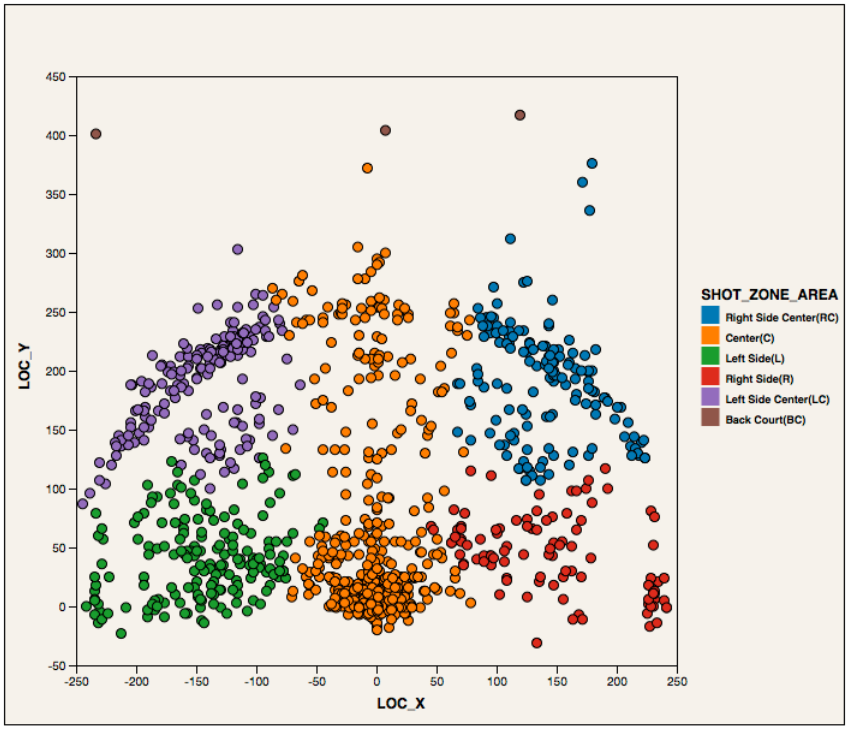
Dataset 4

What sport are we talking about?

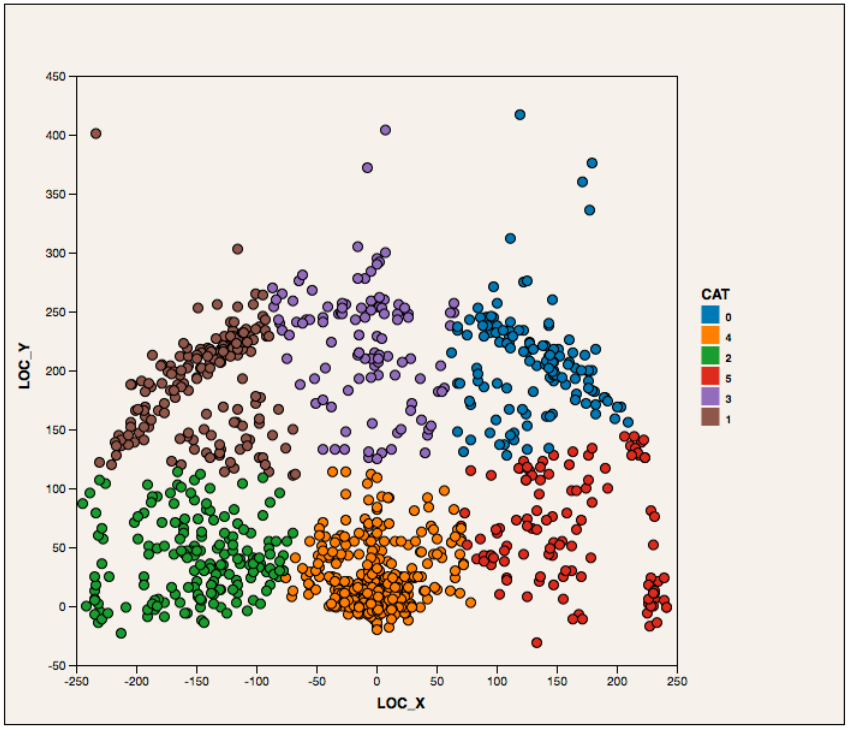


Dataset 4

Original data



Clustered data



Dataset 4

- Let us compare the following adapted key performance indicators (KPI) of two players:

KPI	Player 1	Player 2
FG%	50%	47%
FGA	18	17
MIN	31	30
GP	80	78
FGA LMP		
FG% LMP		
FG% LC		
FG% RC		

Dataset 4

- Let us compare the following adapted key performance indicators (KPI) of two players:

KPI	Player 1	Player 2
FG%	50%	47%
FGA	18	17
MIN	31	30
GP	80	78
FGA LMP	2	7
FG% LMP	6%	24%
FG% LC	41%	48%
FG% RC	38%	57%

- Player 1 is an average shooter that takes easy shots.
- Player 2 is a good shooter that takes difficult shots.

Dataset 5

US airline performance dataset
Cross-filtering and interactive dashboard

This dataset and the associated scripts (html and js) are available at the following github link:

`https://github.com/joelguerrero/joelguerrero.github.io/tree/master/p1/`

The main github.io page is:

`http://joelguerrero.github.io/p1/`

To run the web-based version go to the following address:

`http://joelguerrero.github.io/p1/dataset5/dashboard`

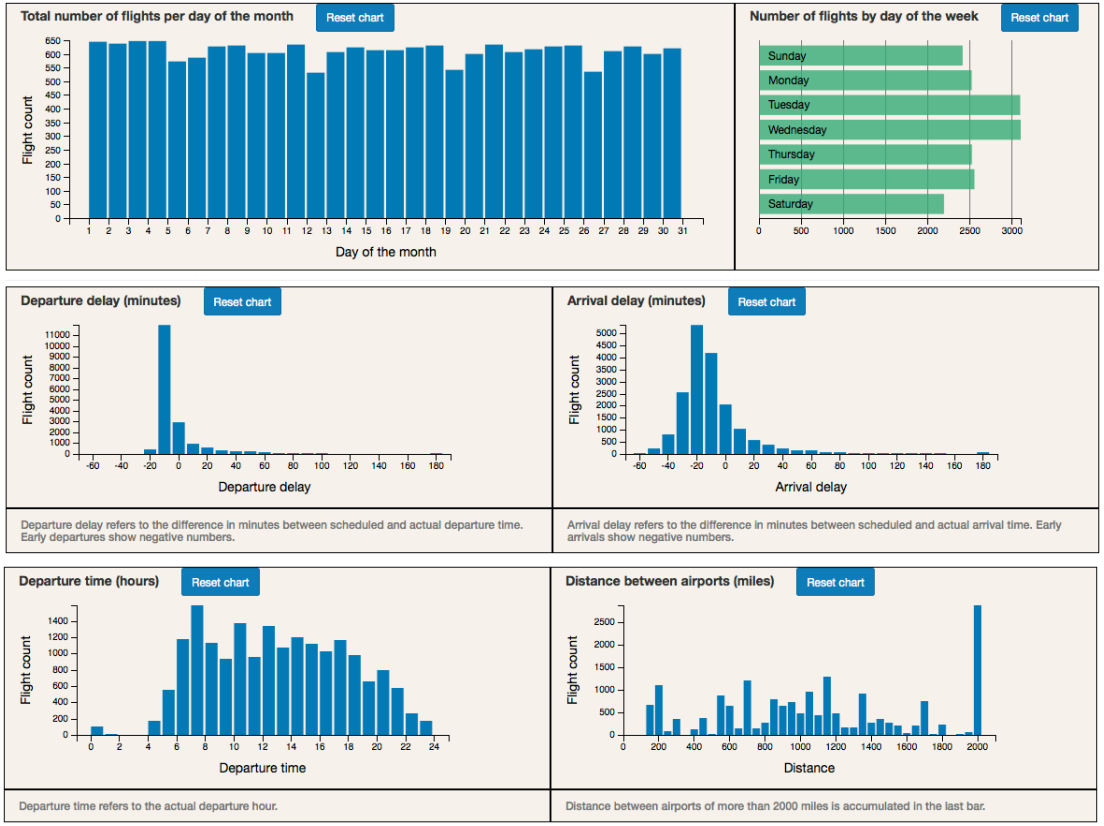
Dataset 5

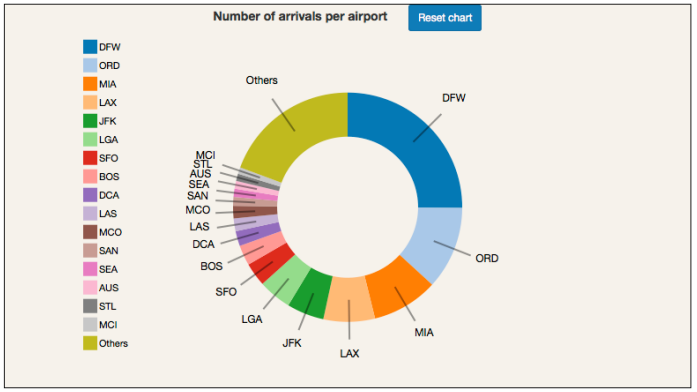
YEAR	MONTH	DAY OF MONTH	DAY OF WEEK	ORIGIN	DESTINATION	DEPARTURE TIME	DEPARTURE DELAY	ARRIVAL DELAY	DISTANCE
2015	9	1	2	JFK	LAX	0853	-7	-28	2475
2015	9	2	3	JFK	LAX	0854	-6	14	2475
2015	9	3	4	JFK	LAX	0857	-3	-14	2475
2015	9	4	5	JFK	LAX	0852	-8	-32	2475
2015	9	5	6	JFK	LAX	0846	-14	-26	2475
2015	9	6	7	JFK	LAX	0855	-5	-58	2475
…	…	…	…	…	…	…	…	…	…
N	N	N	N	N	N	N	N	N	N

- Data obtained from the web.
- Sample size: 10 X 464947
- Numerical data, categorical data and timestamps.
- Tidy data with missing values.

Dataset 5

- This example is about cross filtering data.
- X-filtering is about finding common dimensions, grouping data, using aggregators, filtering data and building interactive dashboards.





Key takeaways

- Data is being used by everybody all the time. The way to analyze and visualize the data is same.
- Data visualization is a powerful way to simplify the complexity in our data and present it in a form which is comprehensible, insightful and actionable.
- Interactive data manipulation and cross-filtering opens a new door in the way we perceive and interpret data.
- Interactive data visualization is an amazing tool for collaboration.
- Data visualization and visual storytelling is not just about a pretty picture but a structured, accurate visual presentation of evidence.

Good intentions

- UI, HCI and UX.
- UQ.
- Advanced ML and SL.
- Big data visualization.

“Principles for the Development of a Complete Mind: Study the science of art. Study the art of science. Develop your senses- especially learn how to see. Realize that everything connects to everything else.”

— Leonardo da Vinci