Lecture Notes for

Machine Learning in Python

Professor Eric Larson

Preprocessing and Visualization

Class Logistics and Agenda

- Participation/Teams
- Be sure you look at **Lab One**!
- Dataset Selection Questions?
- Agenda
  - Finish Pandas Demo with Imputation
  - Data Exploration
  - Data Preprocessing
  - Data Visualization

Last Time

- Datatypes
- Imputation
- Document Features

Feature Type Representation Review

	Attribute	Representation Transformation	Comments
Discrete	Nominal	Any permutation of values **one hot encoding**	If all employee ID numbers were reassigned, would it make any difference?
	Ordinal	An order preserving change of values, i.e., new_value = f(old_value) where f is a monotonic function **integer**	An attribute encompassing the notion of good, better best can be represented equally well by the values {1, 2, 3} or by {0.5, 1, 10}.
Continuous	Interval	new_value = a * old_value + b where a and b are constants **float**	Thus, the Fahrenheit and Celsius temperature scales differ in terms of where their zero value is and the size of a unit (degree).
	Ratio	new_value = a * old_value **float**	Length can be measured in meters or feet.

K-Nearest Neighbors Imputation

TID	Pregnant	BMI	Age	Diabetes
1	Y	33.6	41-50	positive
2	N	26.6	31-40	negative
3	Y	23.3	?	positive
4	?	28.1	21-30	negative
5	N	43.1	31-40	positive
6	Y	25.6	21-30	negative
7	Y	31.0	21-30	positive
8	Y	35.3	?	negative
9	N	30.5	51-60	positive
10	Y	37.6	51-60	positive

For K=3, find 3 closest neighbors

TID	Pregnant	BMI	Age	Diabetes	Dist
3	Y	23.3	?	positive	0
6	Y	25.6	21-30	negative	(0+2.3+1)/3
2	N	26.6	31-40	negative	(1+3.3+1)/3
4	?	28.1	21-30	negative	(4.8+1)/2

How to calculate distance?
- Difference for valid features only
- May need to normalize ranges
- Or weight neighbors differently
- Or have min # of valid features
- Euclidean, city-block, etc.

Feature Hashing

what happens when we get more words?

ID	Slee	Fat	We	Ras	First	Sigh	Why	Fox	Bro	Lazy	Dog	Etc	Stev
	0	1	0	0	2	0	0	0	0	1	0	2	0
	1	1	0	0	1	1	0	0	4	0	1	3	0
	1	1	0	2	1	1	1	0	1	0	0	1	0

or we could have a hashing function, h(x) = y

TID	h(x)=1	h(x)=2	h(x)=3	h(x)=4	h(x)=5	h(x)=6
1	0	1	0	1	2	0
2	1	1	4	0	2	1
3	2	1	1	2	1	1

multiple words mapped to one feature
(want to minimize collisions or make collisions meaningful)

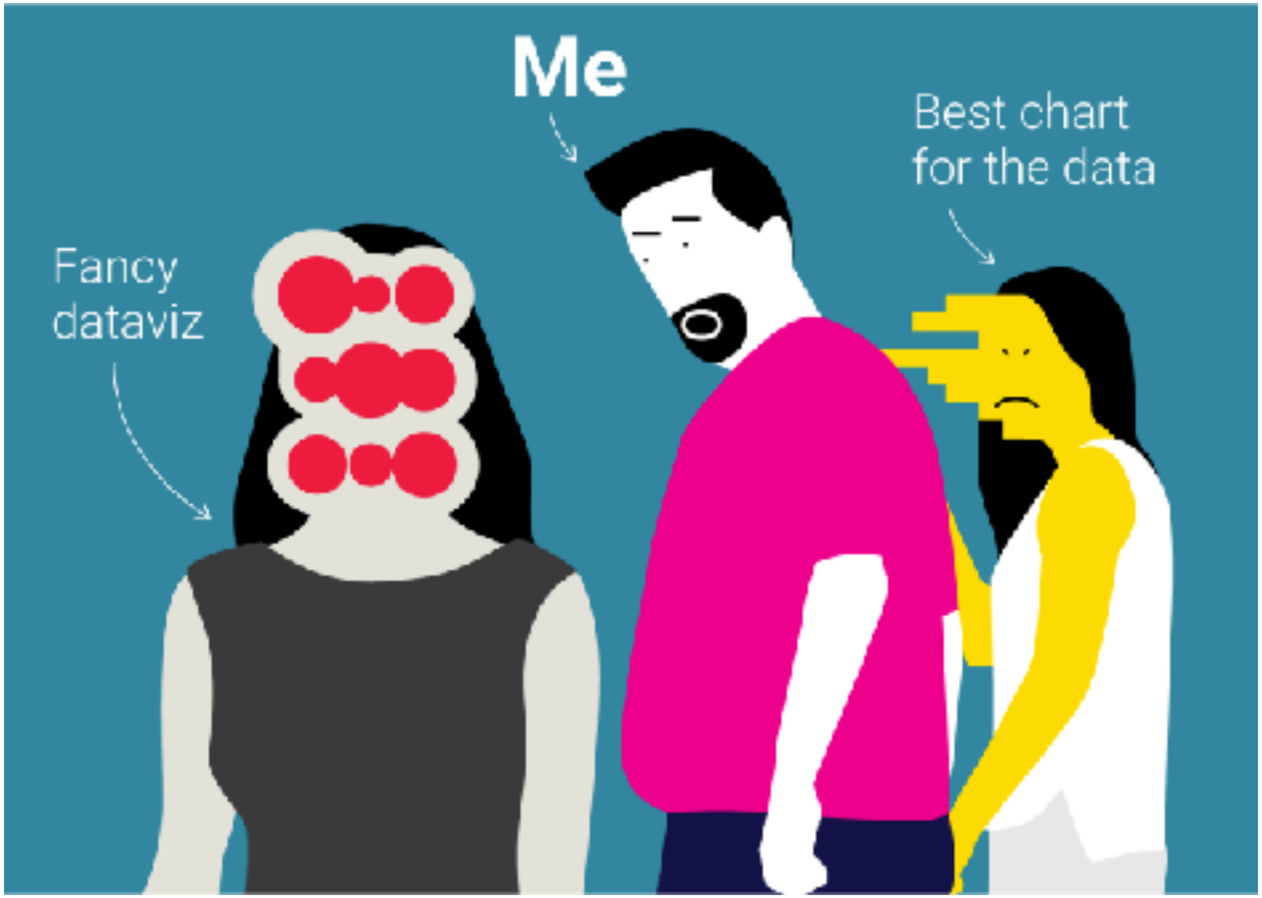
Demo

Start
Pandas demo

- DataFrames
- Loading
- Indexing
- Imputing

`03.Data Visualization.ipynb`

Data Exploration



What is data exploration?

A preliminary exploration of the data to better understand its characteristics.

- Help **select** the **right tool** for preprocessing or analysis

- Exploratory Data Analysis (EDA) by Dr. John Tukey:
  - The focus was visualization
  - Clustering and anomaly detection were viewed as exploratory techniques

- In our discussion,
  - Summary statistics, aggregations
  - Visualizing summaries

Tan, Steinbeck, Kumar

Summary Statistics

- frequency, location, and spread
  - Examples: location by **mean**
    spread by **standard deviation**
- Most summary statistics can be calculated in a single pass through the data

$$\text{sample} \quad \mathrm{mean}(x) = \overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$\text{sample } \mathrm{median}(x) = \begin{cases} x_{(r+1)} & \text{if } m \text{ is odd, i.e., } m = 2r+1 \\ \frac{1}{2}(x_{(r)} + x_{(r+1)}) & \text{if } m \text{ is even, i.e., } m = 2r \end{cases}$$

- For nominal data, mode or frequency is most common

Tan, Steinbeck, Kumar

Measures of Spread

- **Range** is the difference between the max and min
- The **variance** or standard deviation is the most common measure of the spread of a set of points.

$$\text{sample variance}(x) = s_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \overline{x})^2$$

- However, this is also sensitive to outliers, so that other measures are often used.

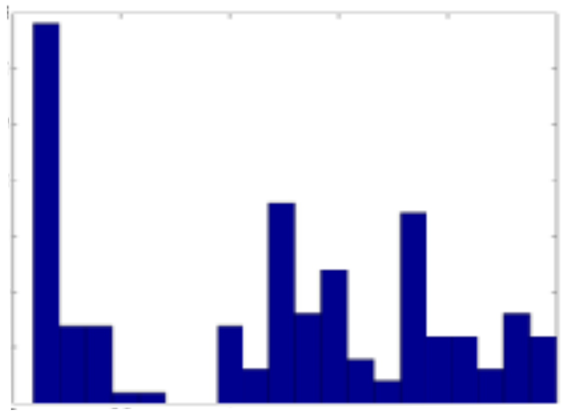
$$\text{AAD}(x) = \frac{1}{m}\sum_{i=1}^{m}|x_i - \overline{x}|$$

$$\text{MAD}(x) = median\Big(\{|x_1 - \overline{x}|, \ldots, |x_m - \overline{x}|\}\Big)$$

$$\text{interquartile range}(x) = x_{75\%} - x_{25\%}$$

Self Test 2a.1

What measure of **spread** is **most appropriate** for the data in the histogram below?



A) Standard Deviation
B) Interquartile Range
C) Median Absolute Difference
D) None of these

Data Preprocessing

Justin Kiggins @neuromusic · 6/19/18

The most successful people I know

1) don't use Matlab

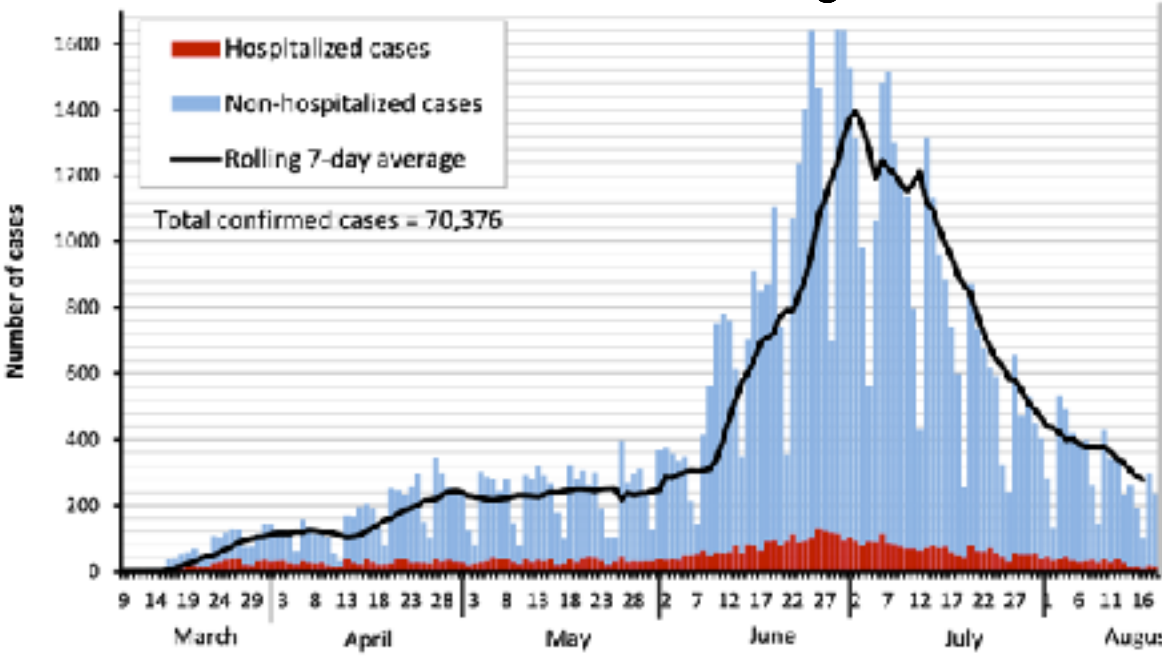
That's... That's pretty much it.

Preprocessing

- Common preprocessing techniques:
  - **Aggregation: Combine features/samples**
    - Reduce the number of attributes or objects
    - Aggregated data tends to be more stable
  - **Transformation: Change of scale**
    - Normalize dynamic ranges
    - More numerically stable when combining
  - **Quantization: Make discrete**
    - More stable
    - More semantically meaningful

Preprocessing: Aggregation

COVID Testing in Dallas County



Steps:

1. map test site to address
2. define address by county
3. group "result" by day w/n county
4. take rolling mean within group
5. repeat for "hospitalized"
6. Sum up bars for total county

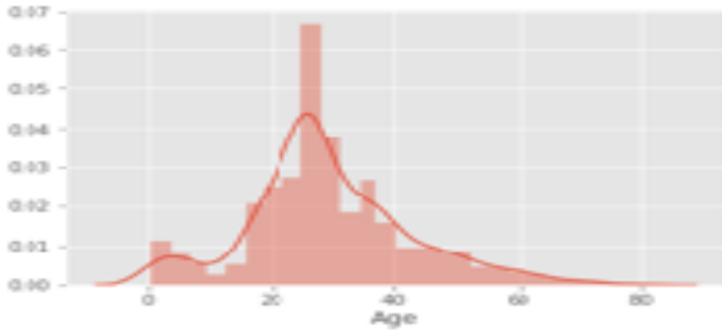
How has aggregation has been used to create these plots?

TID	Location	time	test	Hospitalized
1	test site name	day and hour	test result	yes/no

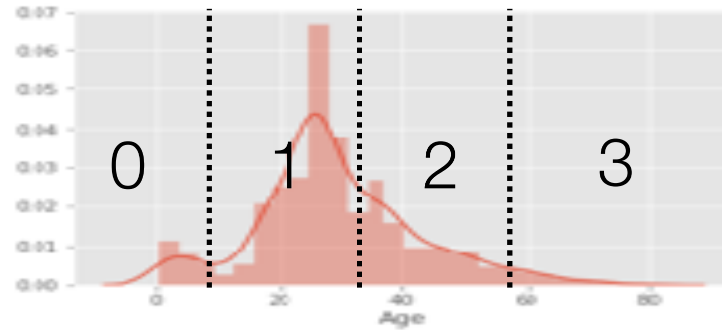
https://www.dallascounty.org/covid-19/

Preprocessing: Quantization

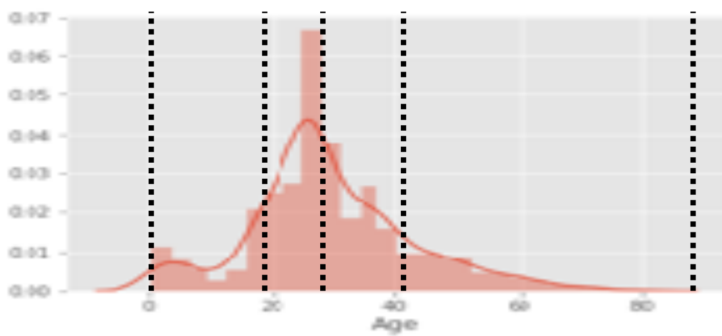
```
pandas.cut(dataframe.var, [5,10,15])
```



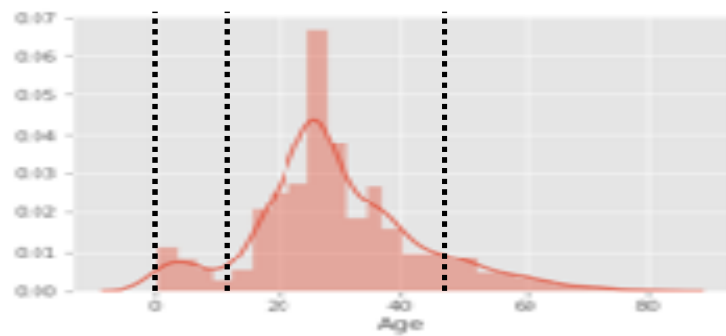
Data



Equal interval width



Equal frequency



clustering: *e.g.*, K-means

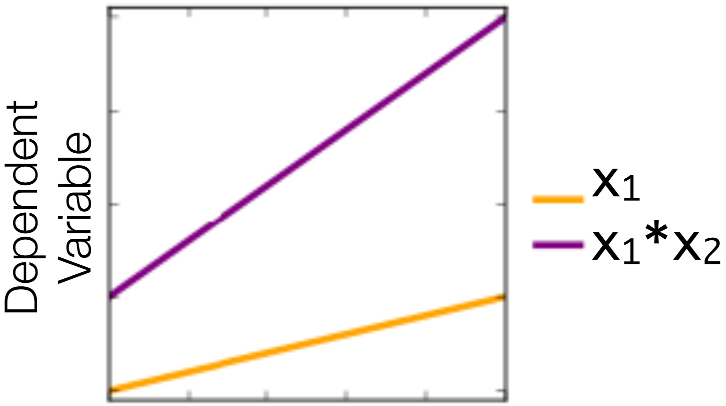
```
num_quantiles = 4
pandas.qcut(dataframe.var, num_quantiles)
```

Preprocessing: Transformation

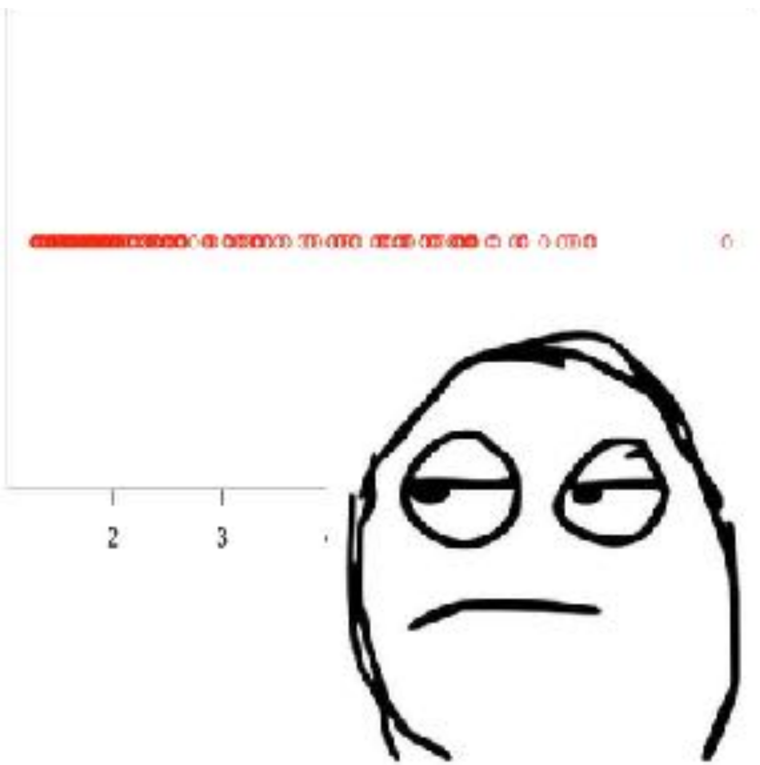
- Monotonically map one set of values to a set of replacement values
  - **Standardization and Normalization**
    - min/max, z-scores

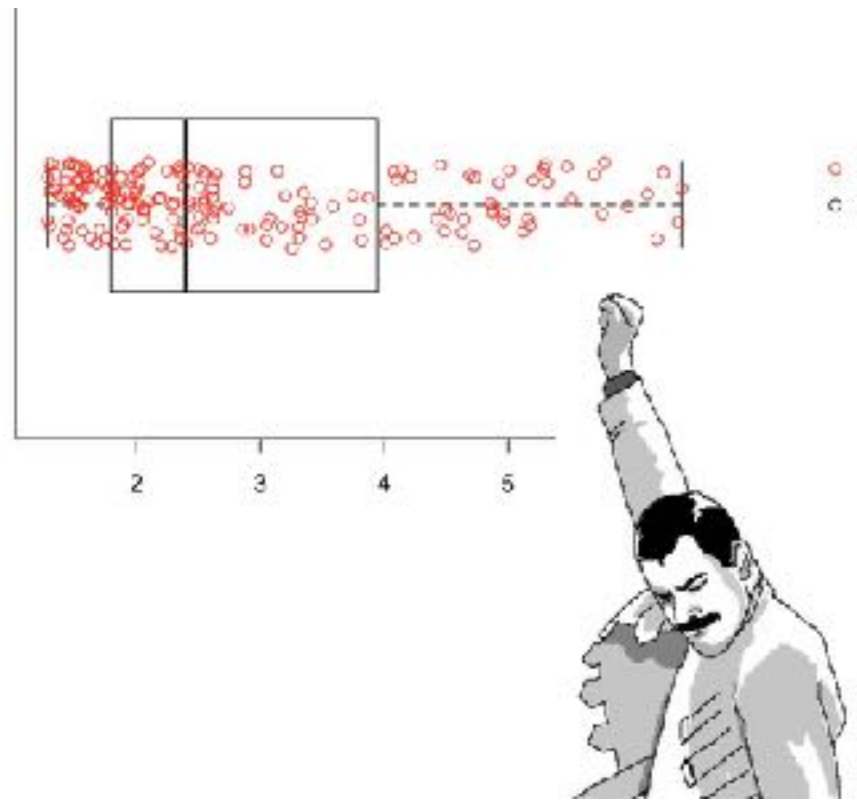
```
df_normalized = (df-df.mean())/(df.std())
```

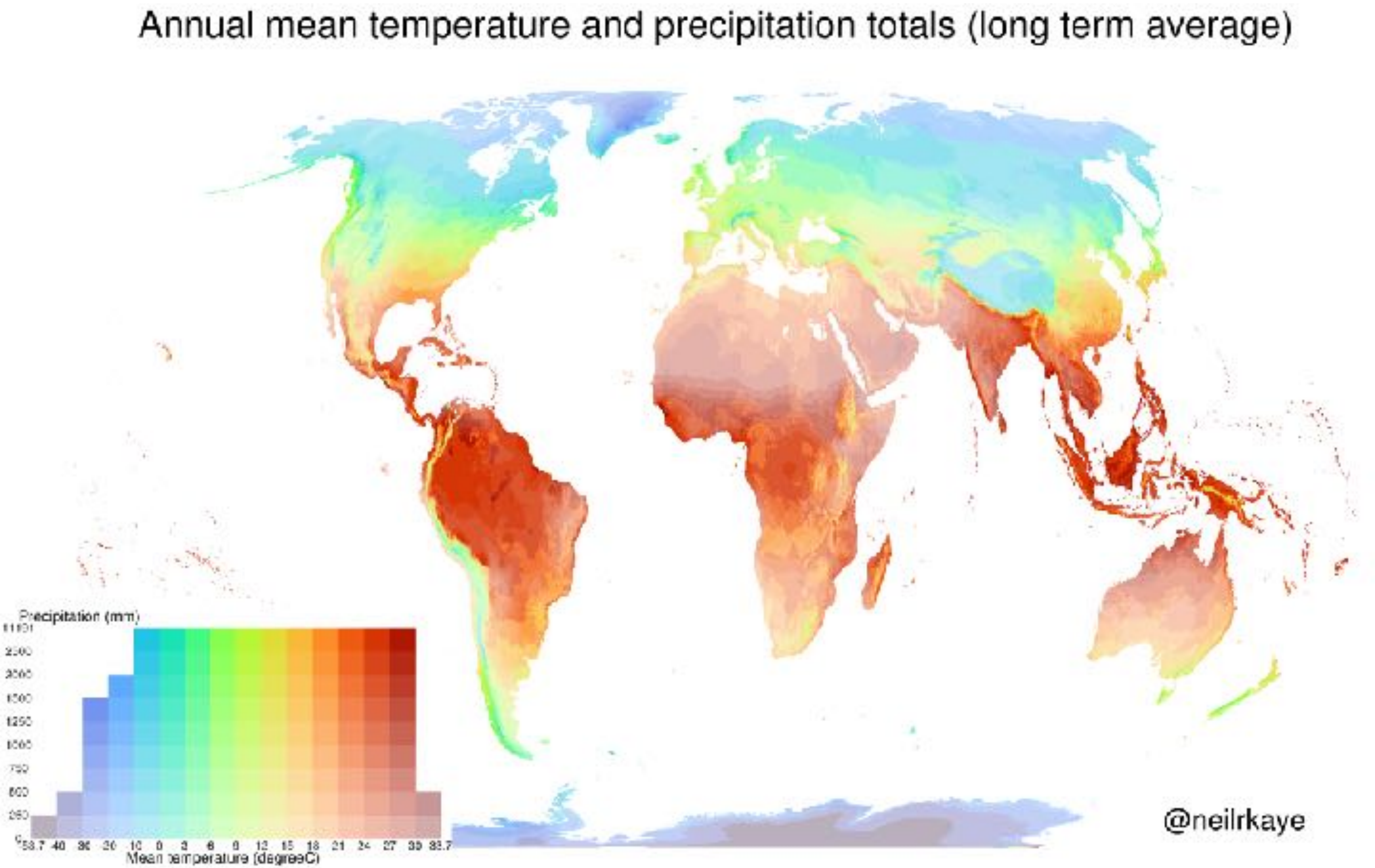
  - **Polynomial and Interaction Variables**
    - `x[:,col=1]**2`
    - `x[:,col=1]*x[:,col=2]`



Data Visualization






Annual mean temperature and precipitation totals (long term average)
Precipitation (mm)
Mean temperature (degreeC)
@neilrkaye

Choosing How/What to Visualize?

- Start with a question you want to understand
- Think about the **best plot** to answer the question
    - Do you have the **right data** for visualizing?
    - Do you need to **worry** about the **amount** of data in the plot (aliasing, low samples, etc.)?
    - Can your question be answered **reliably**?
- **Interpret** the visualization: Did it answer the question?
    - **No**: Think of another visual
    - **Kinda**: Ask a follow up question

Matplotlib

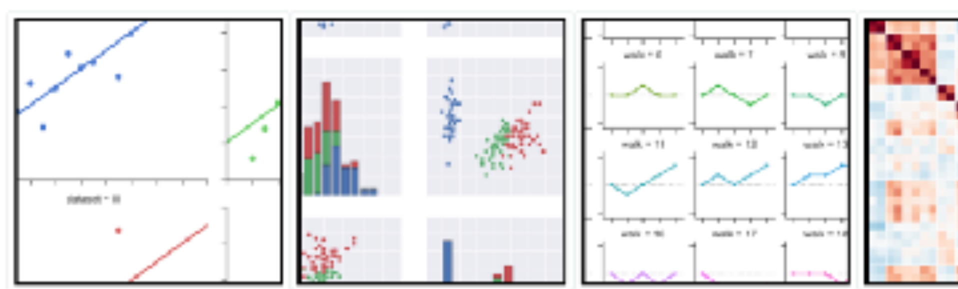
- Python plotting utility
  - Has **low level plotting** functionality
  - Highly **similar to Matlab and R** for plotting
- Extended to be visually more beautiful by
  - **seaborn**: stanford data visualization group

John Hunter (1968-2012)

On August 28 2012, John D. Hunter, the creator of matplotlib, died from complications arising from cancer treatment, after a brief but intense battle with this terrible illness. John is survived by his wife Miriam, his three daughters Rahel, Ava and Clara, his sisters Layne and Mary, and his mother Sarah.

If you have benefited from John's many contributions, please say thanks in the way that would matter most to him. Please consider making a donation to the John Hunter Memorial Fund.

Seaborn: statistical data visualization



Let's look at some graphs

Demo

- You tell me what conclusions we are getting from these graphs
  - Histogram
  - KDE
  - HeatMaps and Correlation
  - Scatter and Scatter Matrix
  - Box / Violin / Swarm

`03.Data Visualization.ipynb`

Matplotlib

Seaborn

Plotly

Let’s look at some graphs

Demo

`03.Data Visualization.ipynb`

Other Tutorials:

https://t.co/zNzD8Q8w5E

http://matplotlib.org/examples/index.html

http://stanford.edu/~mwaskom/software/seaborn/index.html

http://pandas.pydata.org/pandas-docs/stable/visualization.html

http://nbviewer.ipython.org/github/mwaskom/seaborn/blob/master/examples/plotting_distributions.ipynb

For Next Lecture

- Next Time:
  - Finish Visualization Demo
  - First Town Hall Meeting
- Look at chapter 5 of Python Machine Learning

Lecture Notes for

Machine Learning in Python

[👨‍🏫, 👨‍💻, 🐍, 👨‍🔬]

Professor Eric Larson

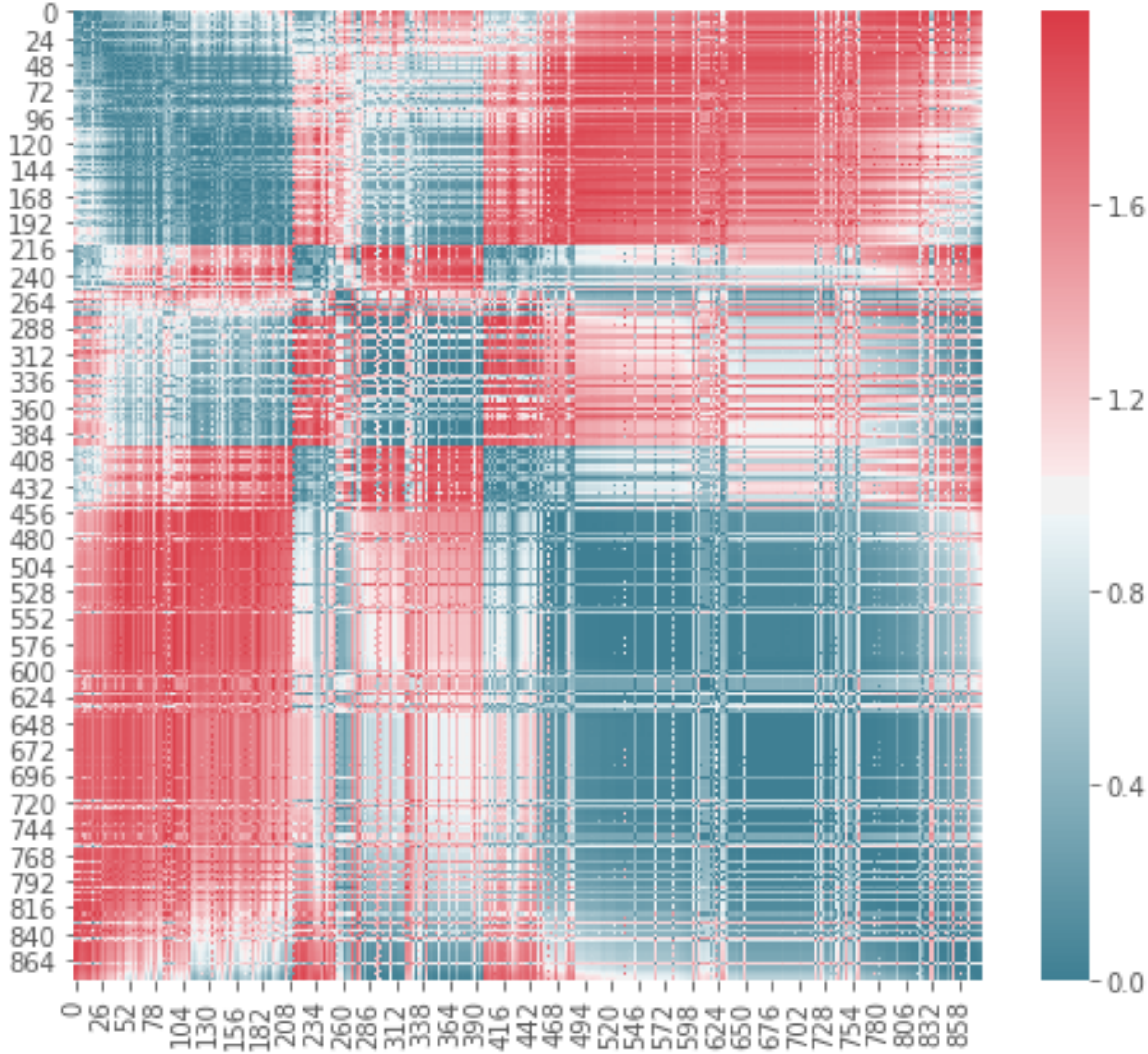
Visualization and Preprocessing

Class Logistics and Agenda

- Finish Visualization Demo
- Town Hall
- Flipped Assignment: Lab One

- Dimensionality Reduction
    - PCA
    - Sampling
    - Kernel Methods
  - Images

What is this plot?



Let’s look at some graphs

Demo

- You tell me what conclusions we are getting from these graphs
  - Histogram
  - KDE
  - HeatMaps and Correlation
  - Scatter and Scatter Matrix
  - Box / Violin / Swarm

`03.Data Visualization.ipynb`

Matplotlib

Seaborn

Plotly

Lab One: Town Hall

Supplemental Slides

Visualization Techniques: Contour Plots

- Contour plots
  - Useful when a continuous attribute is measured on a spatial grid
  - They partition the plane into regions of similar values
  - The contour lines that form the boundaries of these regions connect points with equal values
  - The most common example is contour maps of elevation
  - Can also display temperature, rainfall, air pressure, etc.
    - An example for Sea Surface Temperature (SST) is provided on the next slide

Other Visualization Techniques

- Star Plots
  - Similar approach to parallel coordinates, but axes radiate from a central point
  - The line connecting the values of an object is a polygon
- Chernoff Faces
  - Approach created by Herman Chernoff
  - This approach associates each attribute with a characteristic of a face
  - The values of each attribute determine the appearance of the corresponding facial characteristic
  - Each object becomes a separate face
  - Relies on human's ability to distinguish faces

Challenges of Data Mining

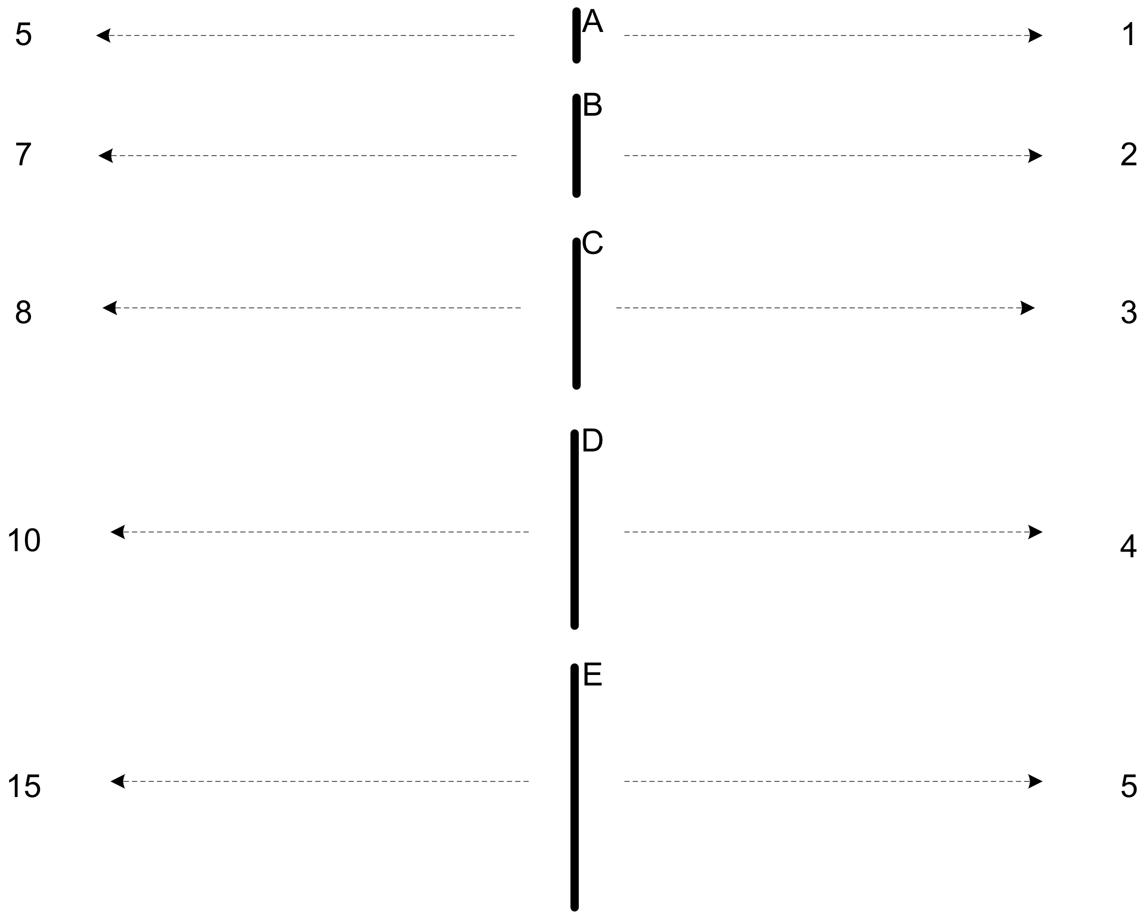
- Scalability
- Dimensionality
- Complex and Heterogeneous Data
- Data Quality
- Data Ownership and Distribution
- Privacy Preservation
- Streaming Data

Important Characteristics of Structured Data

- Dimensionality
  - Curse of Dimensionality
- Sparsity
  - Only presence counts
- Resolution
  - Patterns depend on the scale

Measurement of Length

- The way you measure an attribute is somewhat may not match the attributes properties.



Sampling

- **Sampling is the main technique employed for data selection.**
  - **It is often used for both the preliminary investigation of the data and the final data analysis.**

- **Statisticians sample because obtaining the entire set of data of interest is too expensive or time consuming.**

- **Sampling is used in data mining because processing the entire set of data of interest is too expensive or time consuming.**

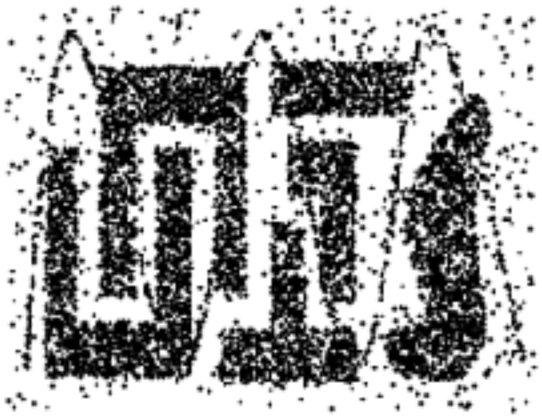
Sampling …

- The key principle for effective sampling is the following:
  - using a sample will work almost as well as using the entire data sets, if the sample is representative
  - A sample is representative if it has approximately the same property (of interest) as the original set of data

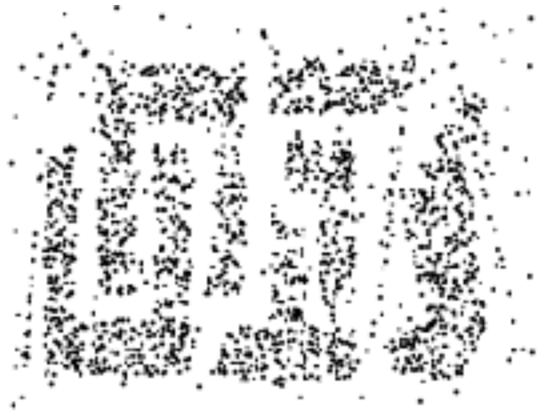
Types of Sampling

- Simple Random Sampling
  - There is an equal probability of selecting any particular item
- Sampling without replacement
  - As each item is selected, it is removed from the population
- Sampling with replacement
  - Objects are not removed from the population as they are selected for the sample.
    - In sampling with replacement, the same object can be picked up more than once
- Stratified sampling
  - Split the data into several partitions; then draw random samples from each partition

Sample Size



8000 points

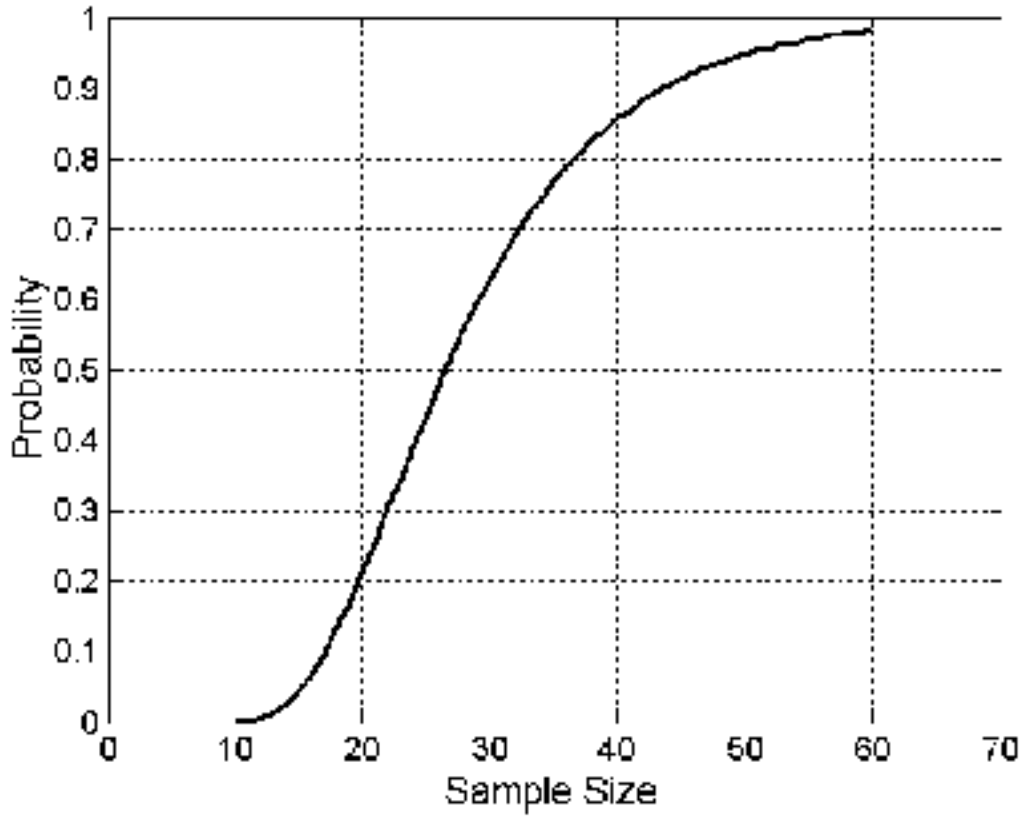


2000 Points

500 Points

Sample Size

- **What sample size is necessary to get at least one object from each of 10 groups.**



Similarity and Dissimilarity

- Similarity
  - Numerical measure of how alike two data objects are.
  - Is higher when objects are more alike.
  - Often falls in the range [0,1]
- Dissimilarity
  - Numerical measure of how different are two data objects
  - Lower when objects are more alike
  - Minimum dissimilarity is often 0
  - Upper limit varies
- Proximity refers to a similarity or dissimilarity

Similarity/Dissimilarity for Simple Attributes

p and *q* are the attribute values for two data objects.

Attribute Type	Dissimilarity	Similarity
Nominal	$d = \begin{cases} 0 & \text{if } p = q \\ 1 & \text{if } p \neq q \end{cases}$	$s = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{if } p \neq q \end{cases}$
Ordinal	$d = \frac{\lvert p-q \rvert}{n-1}$ (values mapped to integers 0 to $n-1$, where n is the number of values)	$s = 1 - \frac{\lvert p-q \rvert}{n-1}$
Interval or Ratio	$d = \lvert p - q \rvert$	$s = -d$, $s = \frac{1}{1+d}$ or $s = 1 - \frac{d - min\_d}{max\_d - min\_d}$

Table 5.1. Similarity and dissimilarity for simple attributes

Euclidean Distance

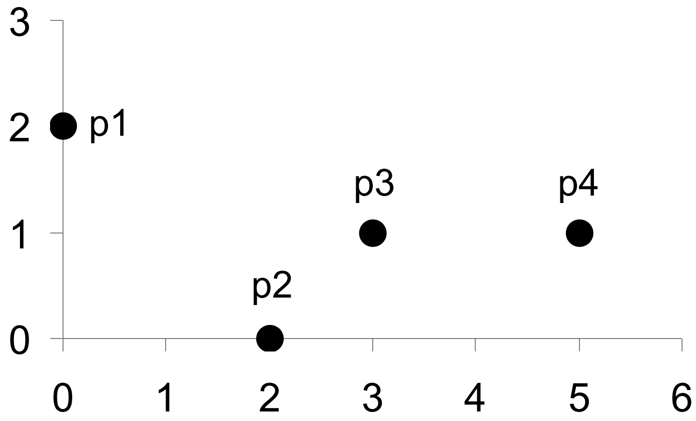
- Euclidean Distance

$$dist = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2}$$

Where n is the number of dimensions (attributes) and p_k and q_k are, respectively, the k^{th} attributes (components) or data objects p and q.

- Standardization is necessary, if scales differ.

Euclidean Distance



point	x	y
p1	0	2
p2	2	0
p3	3	1
p4	5	1

	p1	p2	p3	p4
p1	0	2.828	3.162	5.099
p2	2.828	0	1.414	3.162
p3	3.162	1.414	0	2
p4	5.099	3.162	2	0

Distance Matrix

Minkowski Distance

- Minkowski Distance is a generalization of Euclidean Distance

$$dist = \left(\sum_{k=1}^{n} | p_k - q_k |^r \right)^{\frac{1}{r}}$$

Where r is a parameter, n is the number of dimensions (attributes) and p_k and q_k are, respectively, the kth attributes (components) or data objects p and q.

Minkowski Distance: Examples

- $r = 1$. City block (Manhattan, taxicab, L_1 norm) distance.
  - A common example of this is the Hamming distance, which is just the number of bits that are different between two binary vectors
- $r = 2$. Euclidean distance
- $r \rightarrow \infty$. "supremum" (L_{max} norm, L_∞ norm) distance.
  - This is the maximum difference between any component of the vectors
- Do not confuse *r* with *n*, i.e., all these distances are defined for all numbers of dimensions.

Minkowski Distance

point	x	y
p1	0	2
p2	2	0
p3	3	1
p4	5	1

L1	p1	p2	p3	p4
p1	0	4	4	6
p2	4	0	2	4
p3	4	2	0	2
p4	6	4	2	0

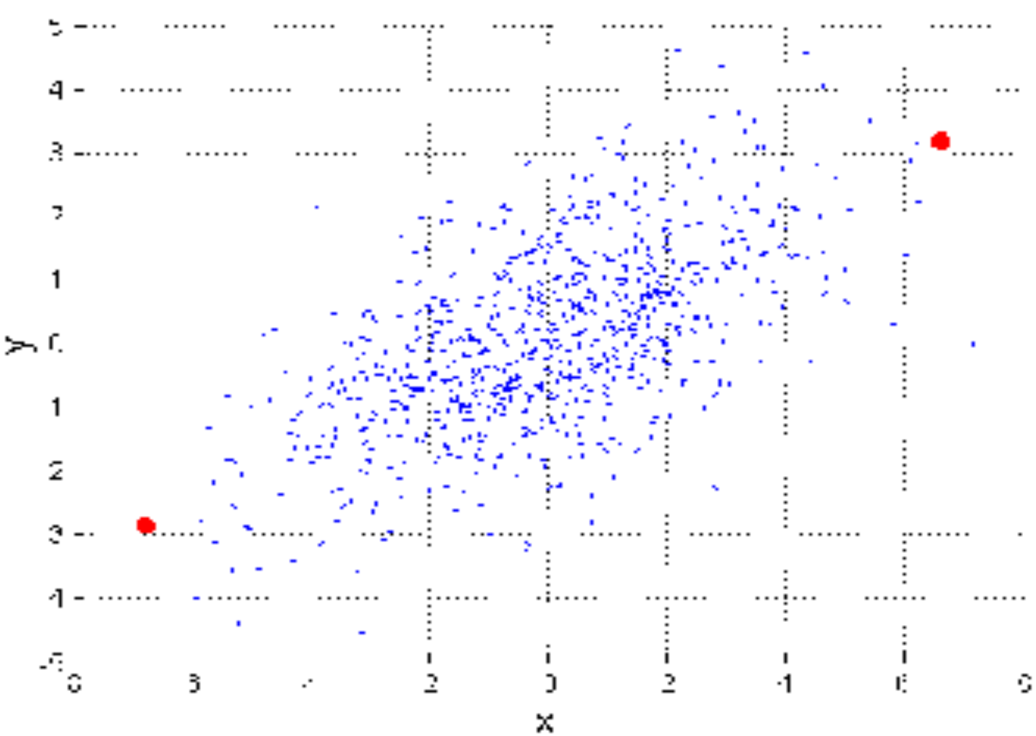
L2	p1	p2	p3	p4
p1	0	2.828	3.162	5.099
p2	2.828	0	1.414	3.162
p3	3.162	1.414	0	2
p4	5.099	3.162	2	0

L_∞	p1	p2	p3	p4
p1	0	2	3	5
p2	2	0	1	3
p3	3	1	0	2
p4	5	3	2	0

Distance Matrix

Mahalanobis Distance

$$mahalanobis(p,q) = (p-q)\sum^{-1}(p-q)^T$$

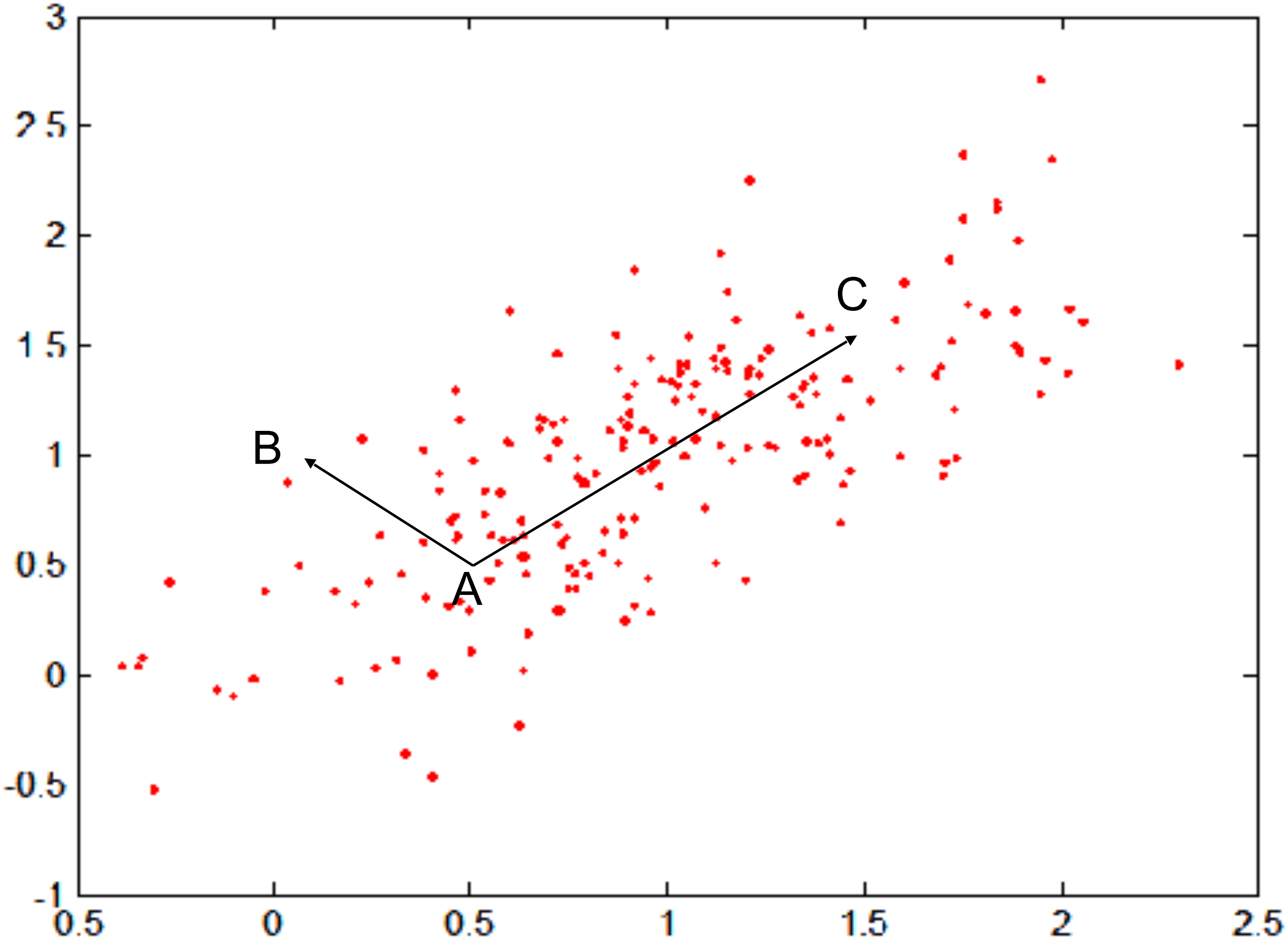


Σ is the covariance matrix of the input data X

$$\Sigma_{j,k} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)(X_{ik}-\overline{X}_k)$$

For red points, the Euclidean distance is 14.7, Mahalanobis distance is 6.

Mahalanobis Distance



Covariance Matrix:

$$\Sigma = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}$$

A: (0.5, 0.5)

B: (0, 1)

C: (1.5, 1.5)

Mahal(A,B) = 5

Mahal(A,C) = 4

Common Properties of a Distance

- Distances, such as the Euclidean distance, have some well known properties.

  1. $d(p, q) \geq 0$ for all p and q and $d(p, q) = 0$ only if $p = q$. (Positive definiteness)
  2. $d(p, q) = d(q, p)$ for all p and q. (Symmetry)
  3. $d(p, r) \leq d(p, q) + d(q, r)$ for all points p, q, and r. (Triangle Inequality)

  where $d(p, q)$ is the distance (dissimilarity) between points (data objects), p and q.

1. A distance that satisfies these properties is a metric

Common Properties of a Similarity

- Similarities, also have some well known properties.

    1. $s(p, q) = 1$ (or maximum similarity) only if $p = q$.
    2. $s(p, q) = s(q, p)$ for all p and q. (Symmetry)

  where $s(p, q)$ is the similarity between points (data objects), p and q.

Similarity Between Binary Vectors

- Common situation is that objects, *p* and *q*, have only binary attributes
- Compute similarities using the following quantities

  M_{01} = the number of attributes where p was 0 and q was 1

  M_{10} = the number of attributes where p was 1 and q was 0

  M_{00} = the number of attributes where p was 0 and q was 0

  M_{11} = the number of attributes where p was 1 and q was 1

- Simple Matching and Jaccard Coefficients

  SMC = number of matches / number of attributes

  = $(M_{11} + M_{00}) / (M_{01} + M_{10} + M_{11} + M_{00})$

  J = number of 11 matches / number of not-both-zero attributes values

  = $(M_{11}) / (M_{01} + M_{10} + M_{11})$

SMC versus Jaccard: Example

$p = $ 1 0 0 0 0 0 0 0 0 0

$q = $ 0 0 0 0 0 0 1 0 0 1

$M_{01} = 2$ (the number of attributes where p was 0 and q was 1)

$M_{10} = 1$ (the number of attributes where p was 1 and q was 0)

$M_{00} = 7$ (the number of attributes where p was 0 and q was 0)

$M_{11} = 0$ (the number of attributes where p was 1 and q was 1)

$$SMC = (M_{11} + M_{00})/(M_{01} + M_{10} + M_{11} + M_{00}) = (0+7) / (2+1+0+7) = 0.7$$

$$J = (M_{11}) / (M_{01} + M_{10} + M_{11}) = 0 / (2 + 1 + 0) = 0$$

Cosine Similarity

- If d_1 and d_2 are two document vectors, then

$$\cos(d_1, d_2) = (d_1 \bullet d_2) / ||d_1||\ ||d_2||,$$

where • indicates vector dot product and $||\ d\ ||$ is the length of vector d.

- Example:

$$d_1 = 3\ 2\ 0\ 5\ 0\ 0\ 0\ 2\ 0\ 0$$
$$d_2 = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 2$$

$$d_1 \bullet d_2 = 3*1 + 2*0 + 0*0 + 5*0 + 0*0 + 0*0 + 0*0 + 2*1 + 0*0 + 0*2 = 5$$

$$||d_1|| = (3*3+2*2+0*0+5*5+0*0+0*0+0*0+2*2+0*0+0*0)^{0.5} = (42)^{0.5} = 6.481$$

$$||d_2|| = (1*1+0*0+0*0+0*0+0*0+0*0+0*0+1*1+0*0+2*2)^{0.5} = (6)^{0.5} = 2.245$$

$$\cos(d_1, d_2) = .3150$$

Extended Jaccard Coefficient (Tanimoto)

- Variation of Jaccard for continuous or count attributes
  - Reduces to Jaccard for binary attributes

$$T(p,q) = \frac{p \bullet q}{|p\|^2 + \|q\|^2 - p \bullet q}$$

Correlation

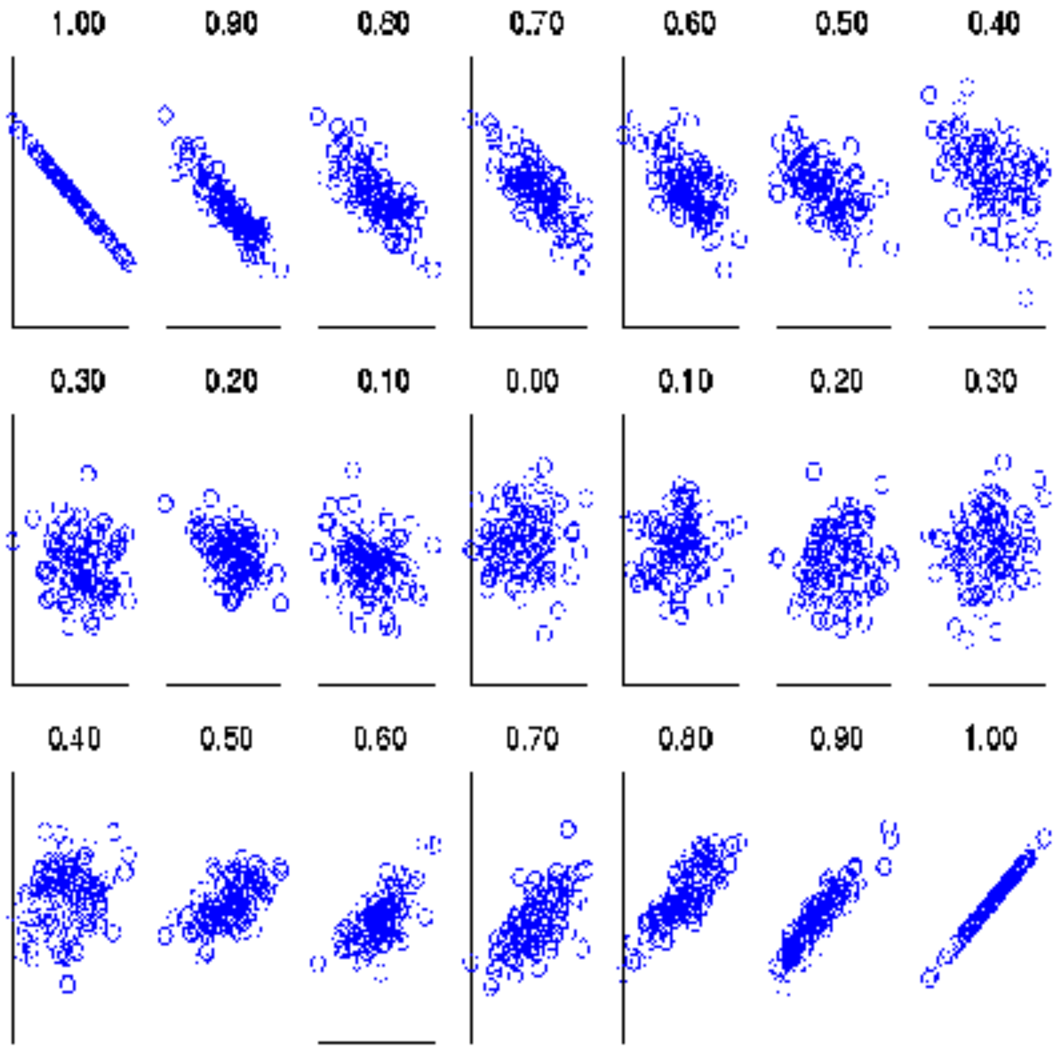
- Correlation measures the linear relationship between objects
- To compute correlation, we standardize data objects, p and q, and then take their dot product

$$p'_k = (p_k - mean(p)) / std(p)$$

$$q'_k = (q_k - mean(q)) / std(q)$$

$$correlation(p, q) = p' \bullet q'$$

Visually Evaluating Correlation



Scatter plots showing the similarity from –1 to 1.

General Approach for Combining Similarities

- Sometimes attributes are of many different types, but an overall similarity is needed.

1. For the k^{th} attribute, compute a similarity, s_k, in the range $[0, 1]$.
2. Define an indicator variable, δ_k, for the k_{th} attribute as follows:

$$\delta_k = \begin{cases} 0 & \text{if the } k^{th} \text{ attribute is a binary asymmetric attribute and both objects have} \\ & \text{a value of } 0 \text{, or if one of the objects has a missing values for the } k^{th} \text{ attribute} \\ 1 & \text{otherwise} \end{cases}$$

3. Compute the overall similarity between the two objects using the following formula:

$$similarity(p, q) = \frac{\sum_{k=1}^{n} \delta_k s_k}{\sum_{k=1}^{n} \delta_k}$$

Using Weights to Combine Similarities

- May not want to treat all attributes the same.
  - Use weights w_k which are between 0 and 1 and sum to 1.

$$similarity(p,q) = \frac{\sum_{k=1}^{n} w_k \delta_k s_k}{\sum_{k=1}^{n} \delta_k}$$

$$distance(p,q) = \left(\sum_{k=1}^{n} w_k |p_k - q_k|^r \right)^{1/r}.$$

Density

- Density-based clustering require a notion of density

- Examples:
    - Euclidean density
        - Euclidean density = number of points per unit volume

    - Probability density

    - Graph-based density

Euclidean Density – Cell-based

- Simplest approach is to divide region into a number of rectangular cells of equal volume and define density as # of points the cell contains

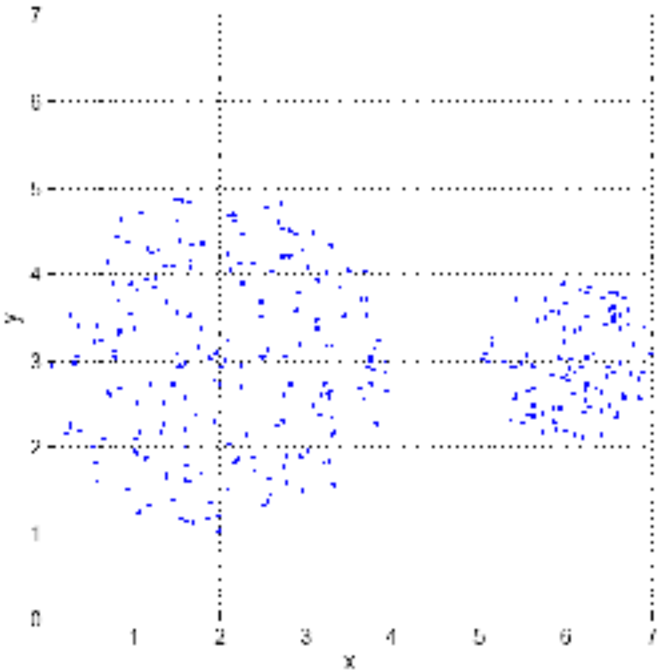


Figure 7.13. Cell-based density.

0	0	0	0	0	0	0
0	0	0	0	0	0	0
4	17	18	6	0	0	0
14	14	13	13	0	18	27
11	18	10	21	0	24	31
3	20	14	4	0	0	0
0	0	0	0	0	0	0

Table 7.6. Point counts for each grid cell.

Euclidean Density – Center-based

- Euclidean density is the number of points within a specified radius of the point

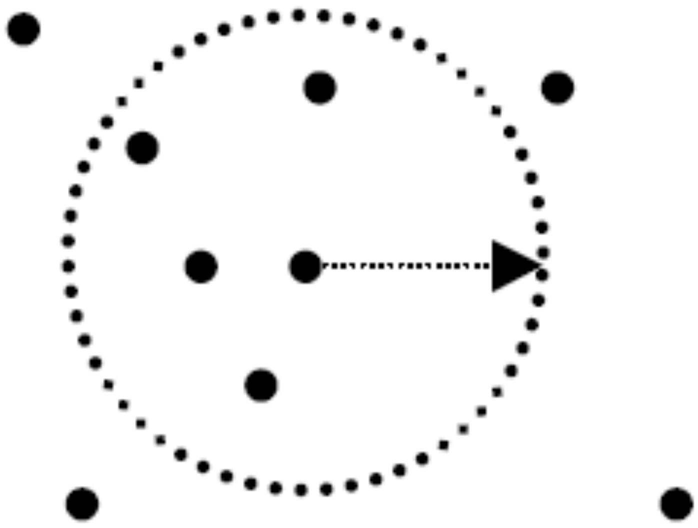


Figure 7.14. Illustration of center-based density.

Feature Subset Selection

- Another way to reduce dimensionality of data

- Redundant features
  - duplicate much or all of the information contained in one or more other attributes
  - Example: purchase price of a product and the amount of sales tax paid

- Irrelevant features
  - contain no information that is useful for the data mining task at hand
  - Example: students' ID is often irrelevant to the task of predicting students' GPA

Feature Subset Selection

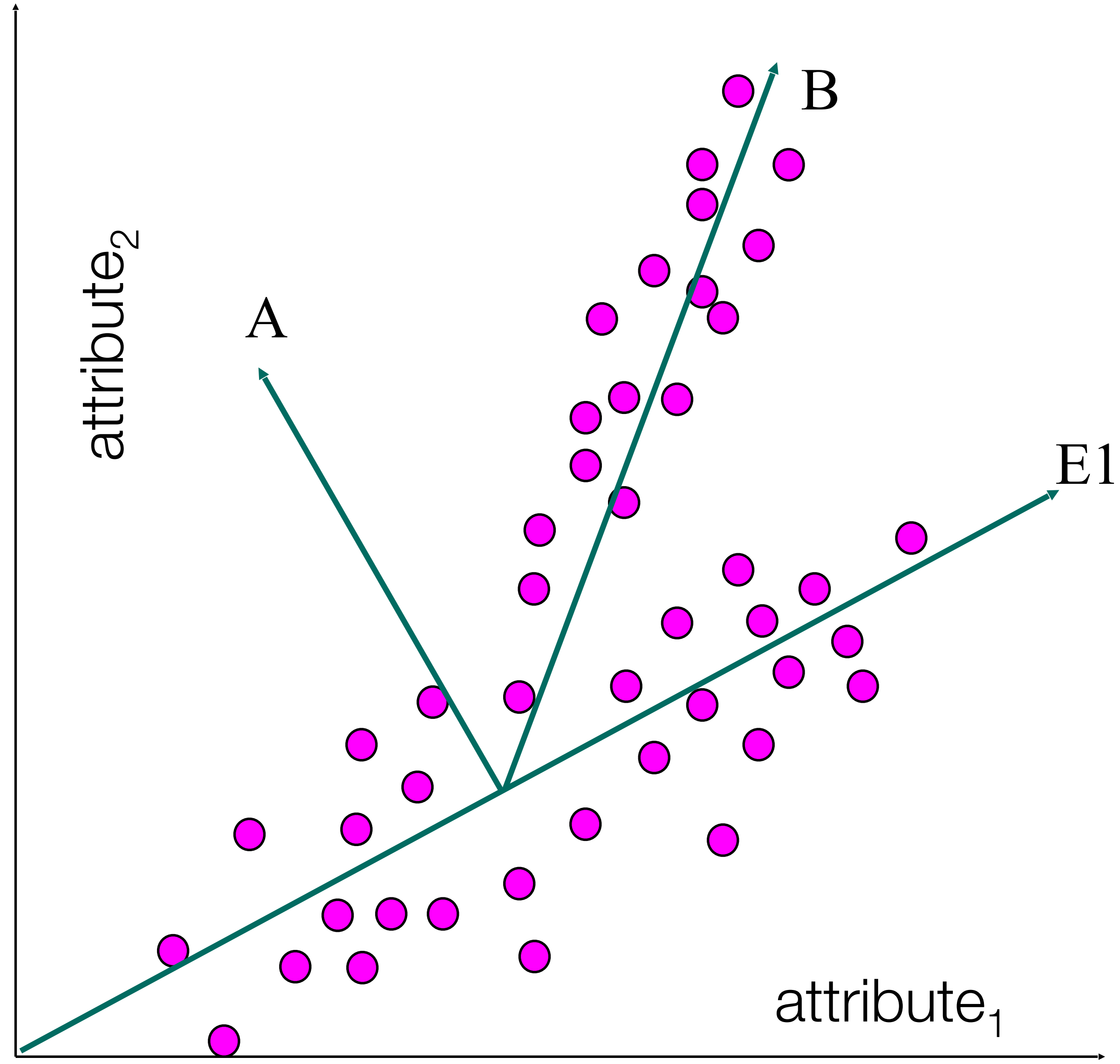
- Techniques:
  - Brute-force approch:
    - Try all possible feature subsets as input to data mining algorithm
  - Embedded approaches:
    - Feature selection occurs naturally as part of the data mining algorithm
  - Filter approaches:
    - Features are selected before data mining algorithm is run
  - Wrapper approaches:
    - Use the data mining algorithm as a black box to find best subset of attributes

Feature Creation

- Create new attributes that can capture the important information in a data set much more efficiently than the original attributes

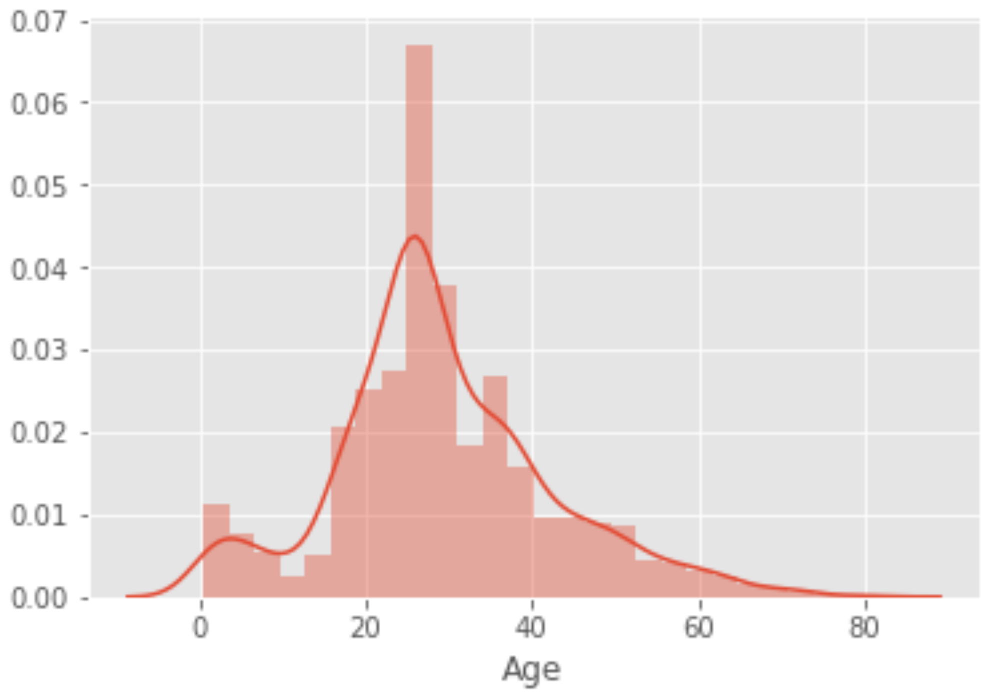
- Three general methodologies:
  - Feature Extraction
    - domain-specific
  - Mapping Data to New Space
  - Feature Construction
    - combining features

Dimensionality Reduction: PCA



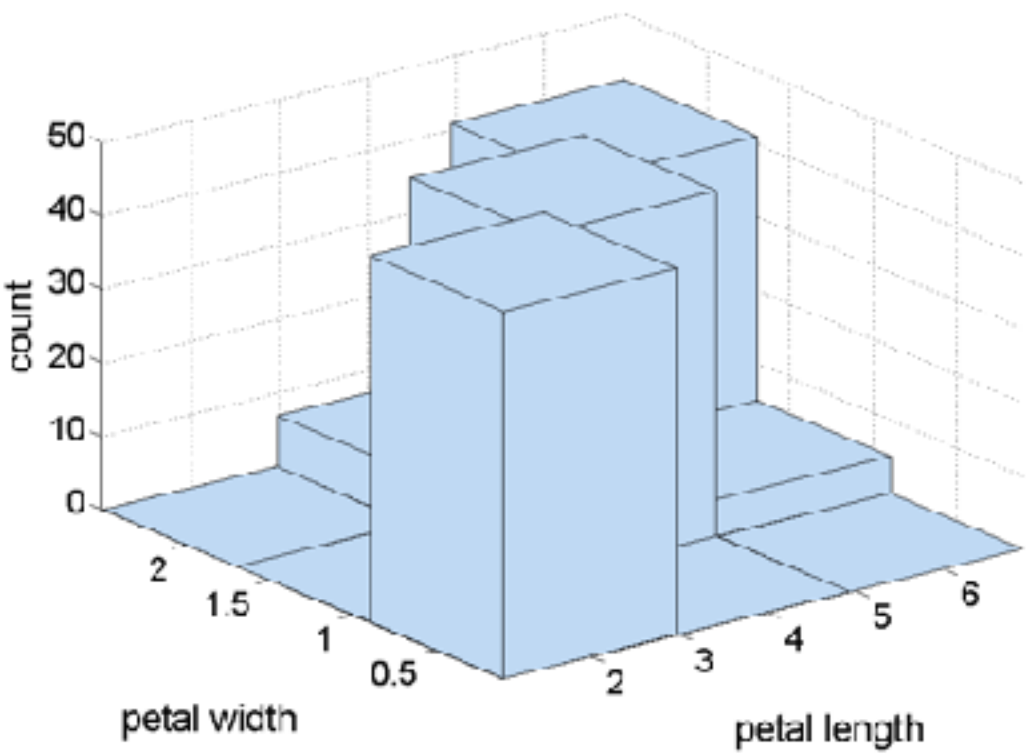
Visualization Techniques: Distributions

- Histogram
  - Usually shows the distribution of values of a single variable
  - Divide the values into bins and show a bar plot of the number of objects in each bin.
- KDE
  - Add up Gaussian underneath each point value
  - STD of gaussian is equivalent to number of bins in histogram

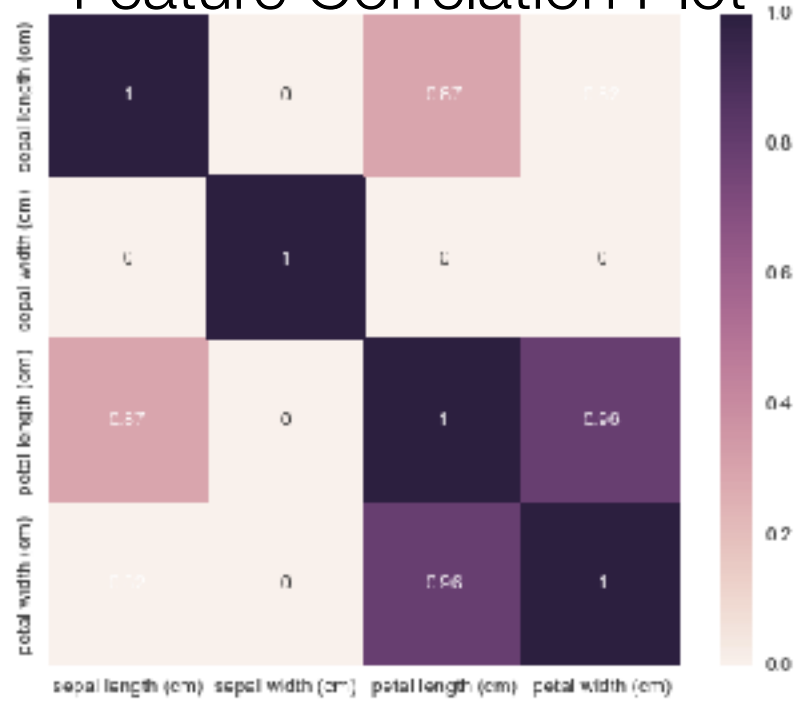


Two-Dimensional Distributions

- Estimate the joint distribution of the values of two attributes
- Example: petal width and petal length
    - What does this tell us?



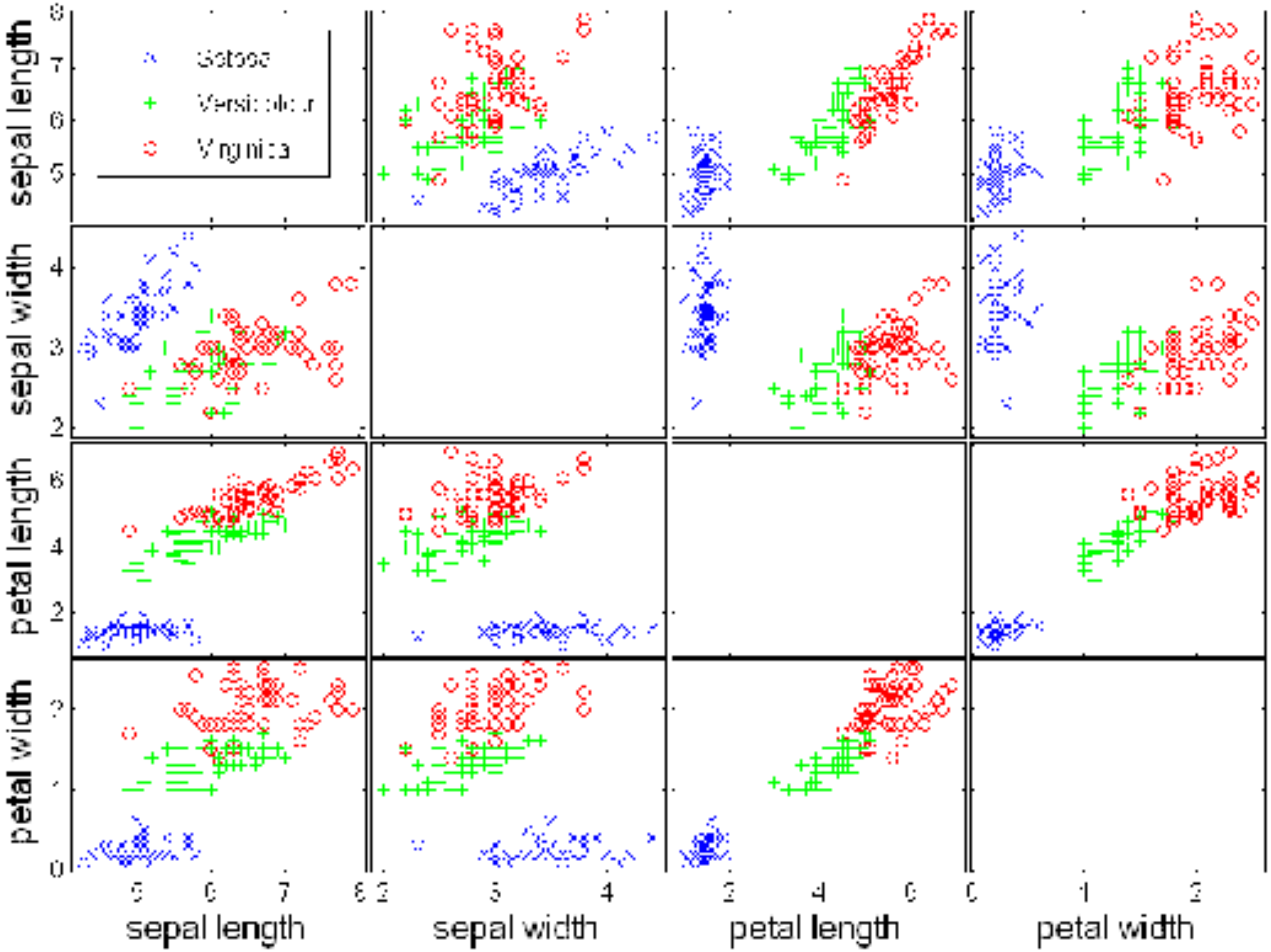
Feature Correlation Plot



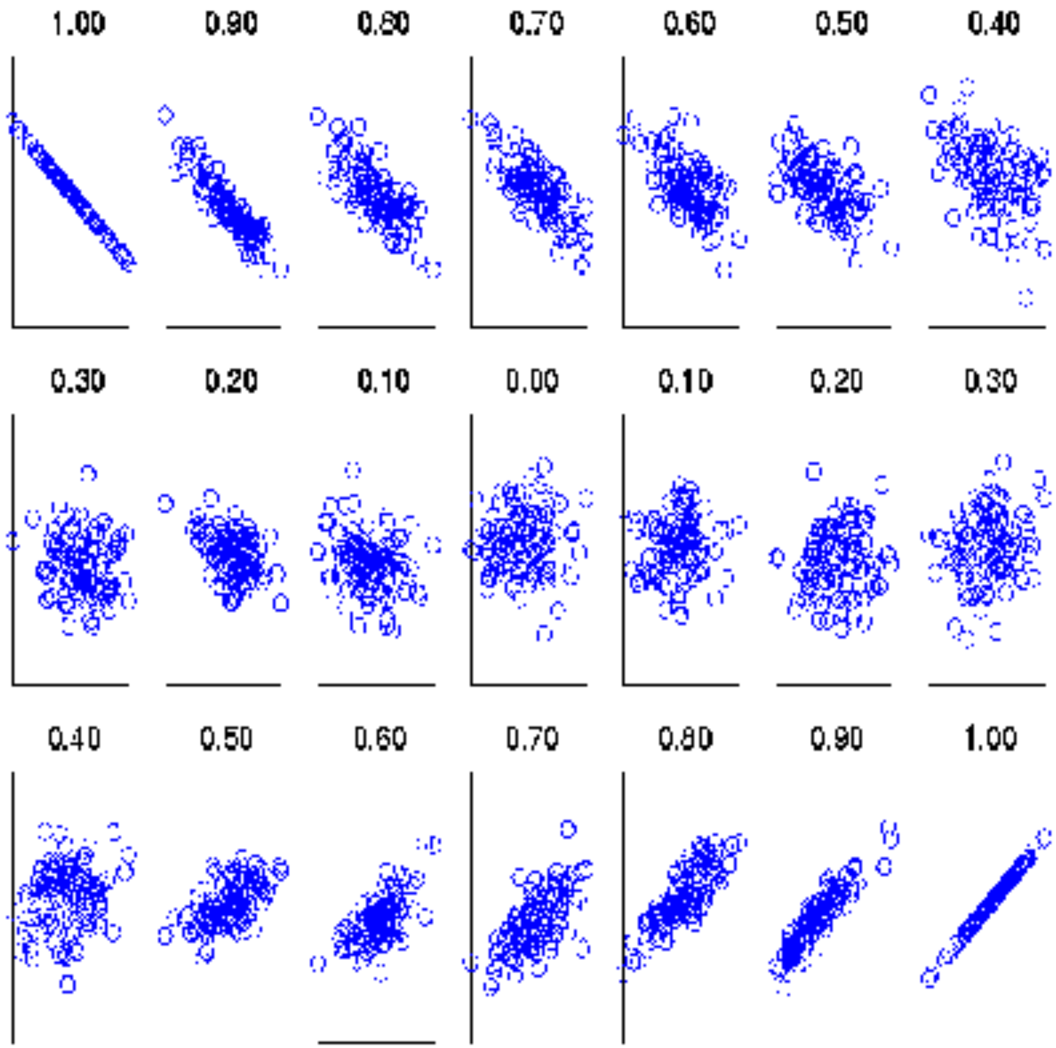
Visualization Techniques: Scatter Plots

- Scatter plots
  - Two-dimensional scatter plots most common
  - **Additional attributes** can be displayed by using the **size, shape, and color** of the markers that represent the objects
  - **Interactivity** can add **insight**
  - It is useful to have **matrices of scatter plots** to compactly summarize the relationships of several pairs of attributes
  - Good for numeric data, but needs **jitter** for categorical data

Scatter Plot Matrix Colored by Class

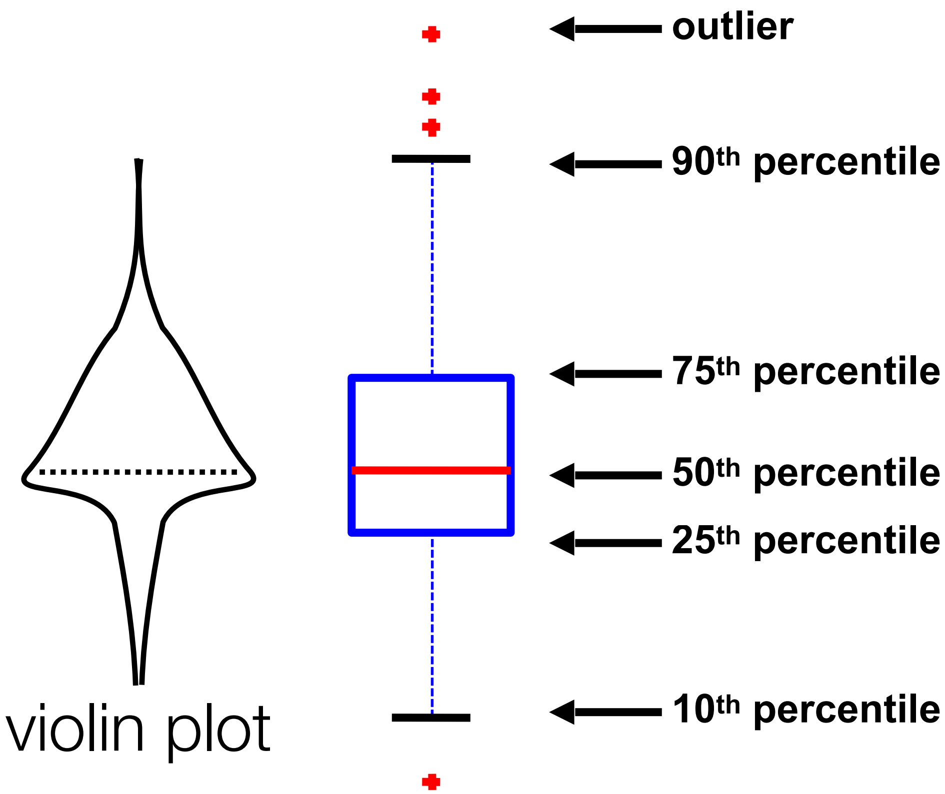


Visually Evaluating Correlation



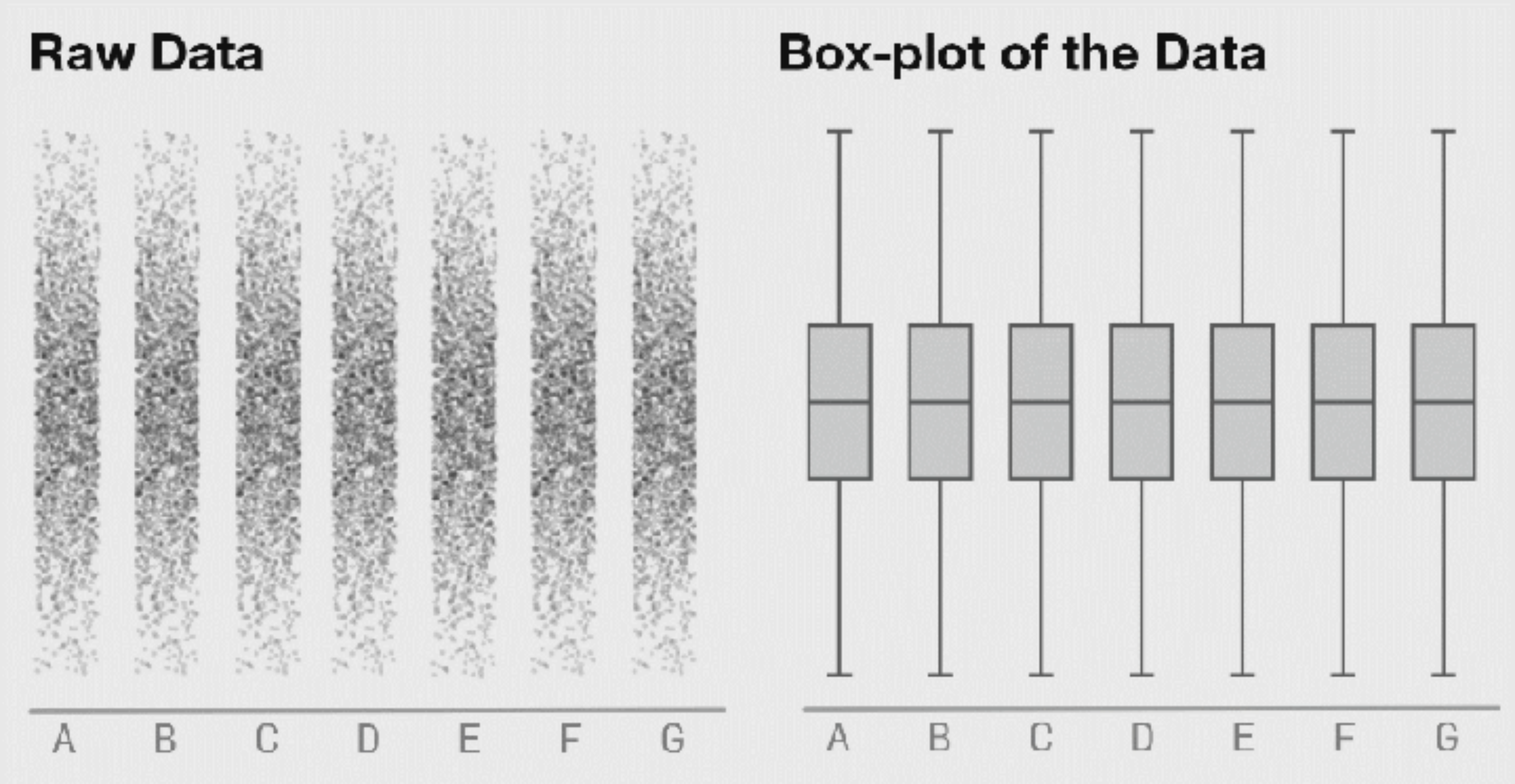
Scatter plots showing the similarity from –1 to 1.

- Box Plots
  - Invented by J. Tukey
  - Another way of displaying the distribution of data
  - Following figure shows the basic part of a box plot



Visualization Techniques: Box Plots

- Box Plots

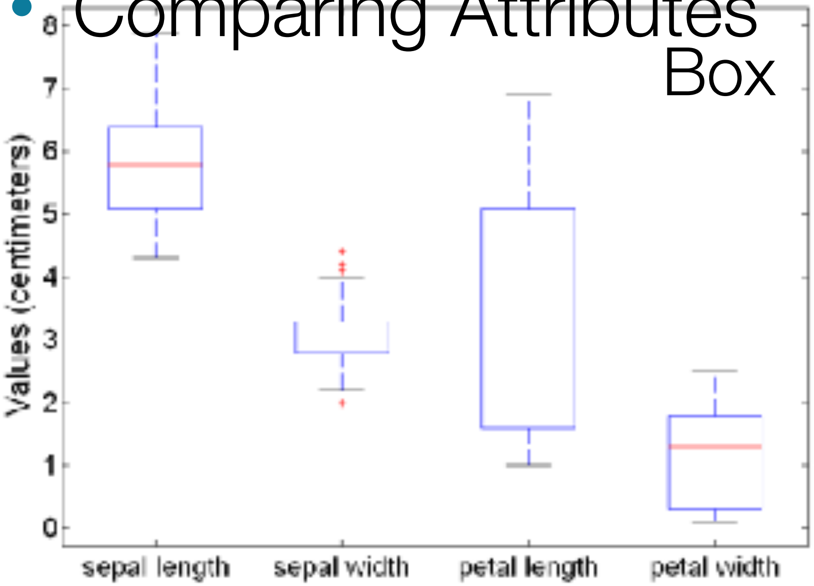


https://www.reddit.com/r/dataisbeautiful/comments/6ssmzc/be_wary_of_boxplots_they_could_be_hiding/

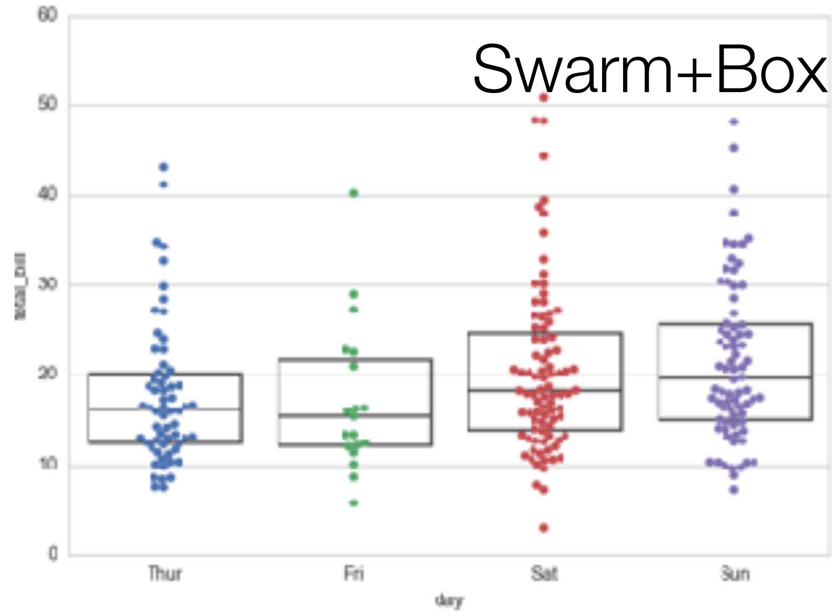
Example: Comparing Attributes

- Comparing Attributes

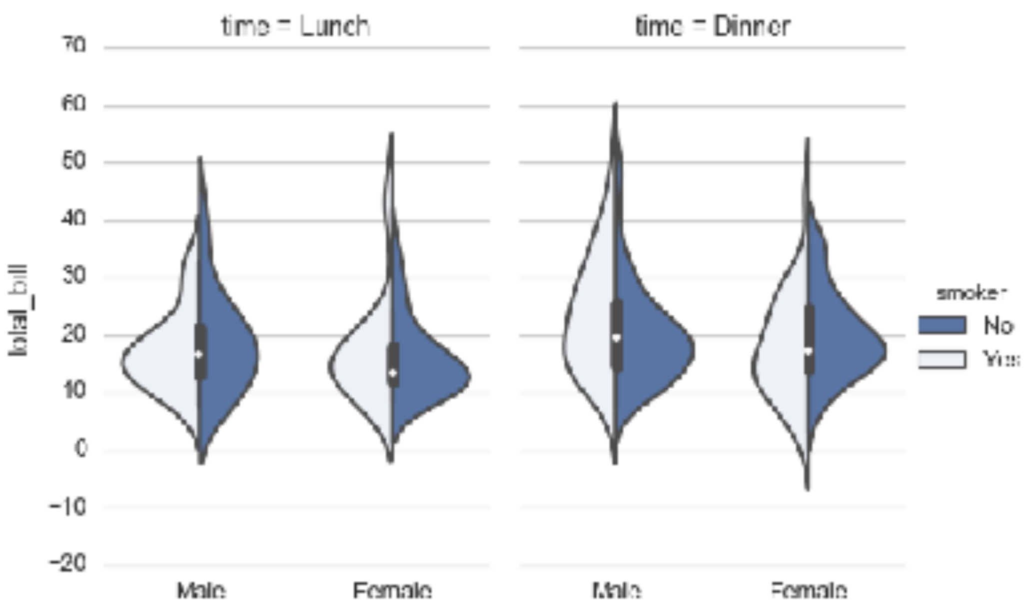
Box



Swarm+Box



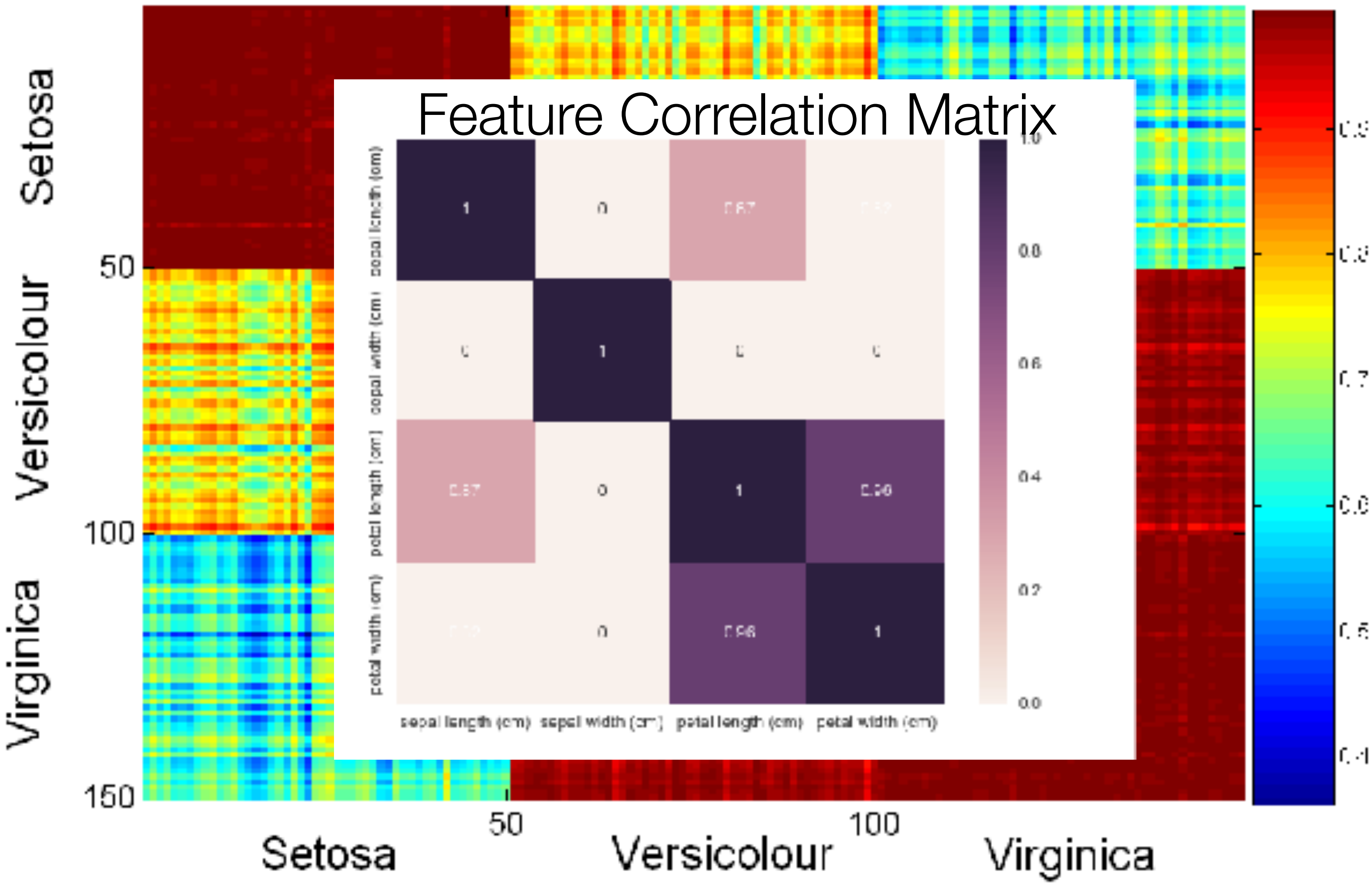
Mixed Violin + Box



Visualization Techniques: Matrix Plots

- Matrix plots (typically heatmaps)
  - Plot some data matrix
  - This can be useful when objects are sorted well
  - Typically, the attributes are normalized to prevent one attribute from dominating the plot
  - Plots of similarity or distance matrices can also be useful for visualizing the relationships between objects

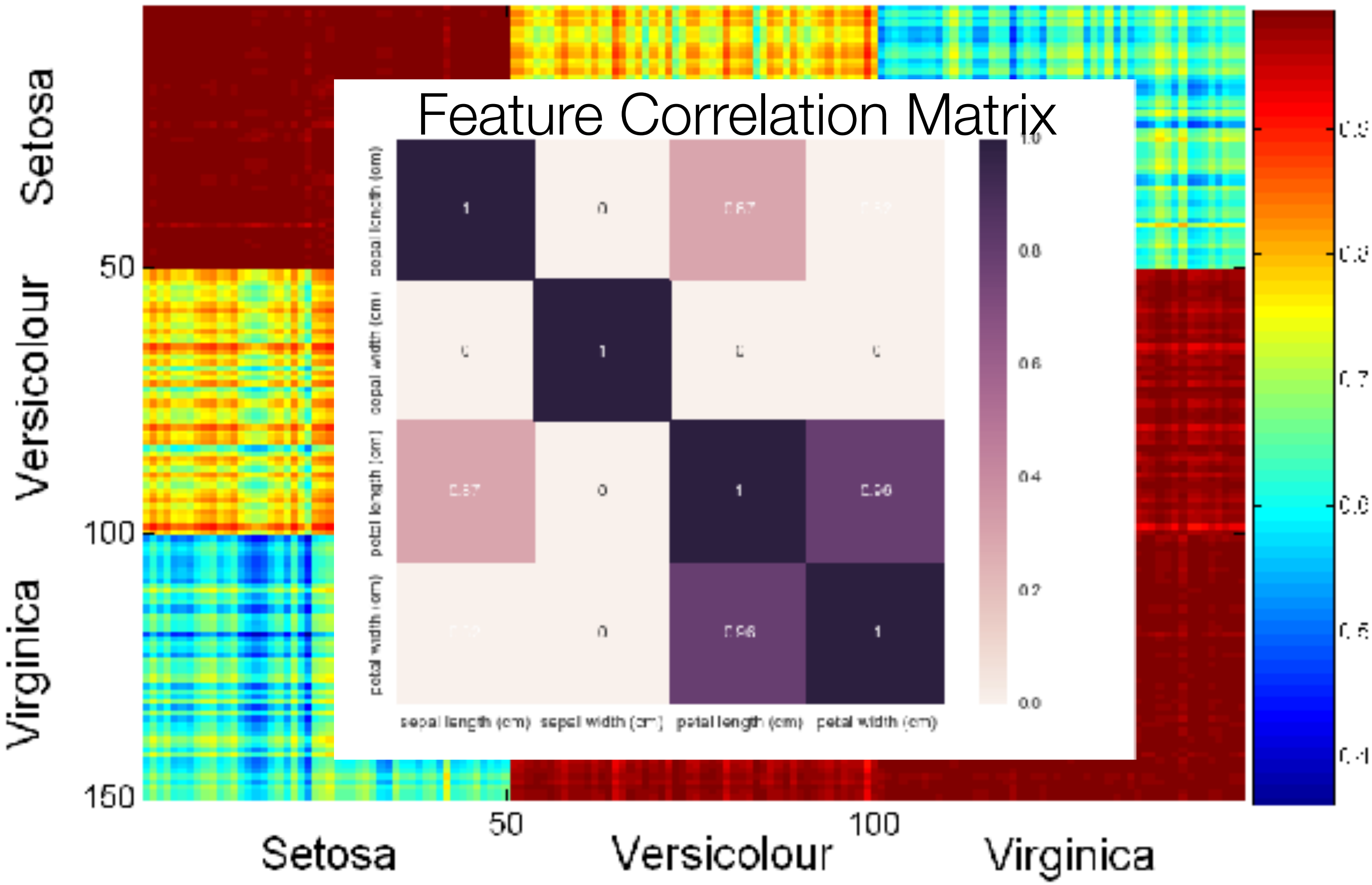
Instance Correlation Matrix



Visualization Techniques: Matrix Plots

- Matrix plots (typically heatmaps)
  - Plot some data matrix
  - This can be useful when objects are sorted well
  - Typically, the attributes are normalized to prevent one attribute from dominating the plot
  - Plots of similarity or distance matrices can also be useful for visualizing the relationships between objects

Instance Correlation Matrix



Visualization Techniques: Parallel Coordinates

- Parallel Coordinates
  - Used to plot the attribute values of multi-dimensional data
  - Instead of using perpendicular axes, use a set of parallel axes
  - The attribute values of each object are plotted as a point on each corresponding coordinate axis and the points are connected by a line
  - Thus, each object is represented as a line
  - Often, the lines representing a distinct class of objects group together, at least for some attributes
  - Ordering of attributes is important in seeing such groupings

Parallel Coordinates Plots for Iris Data