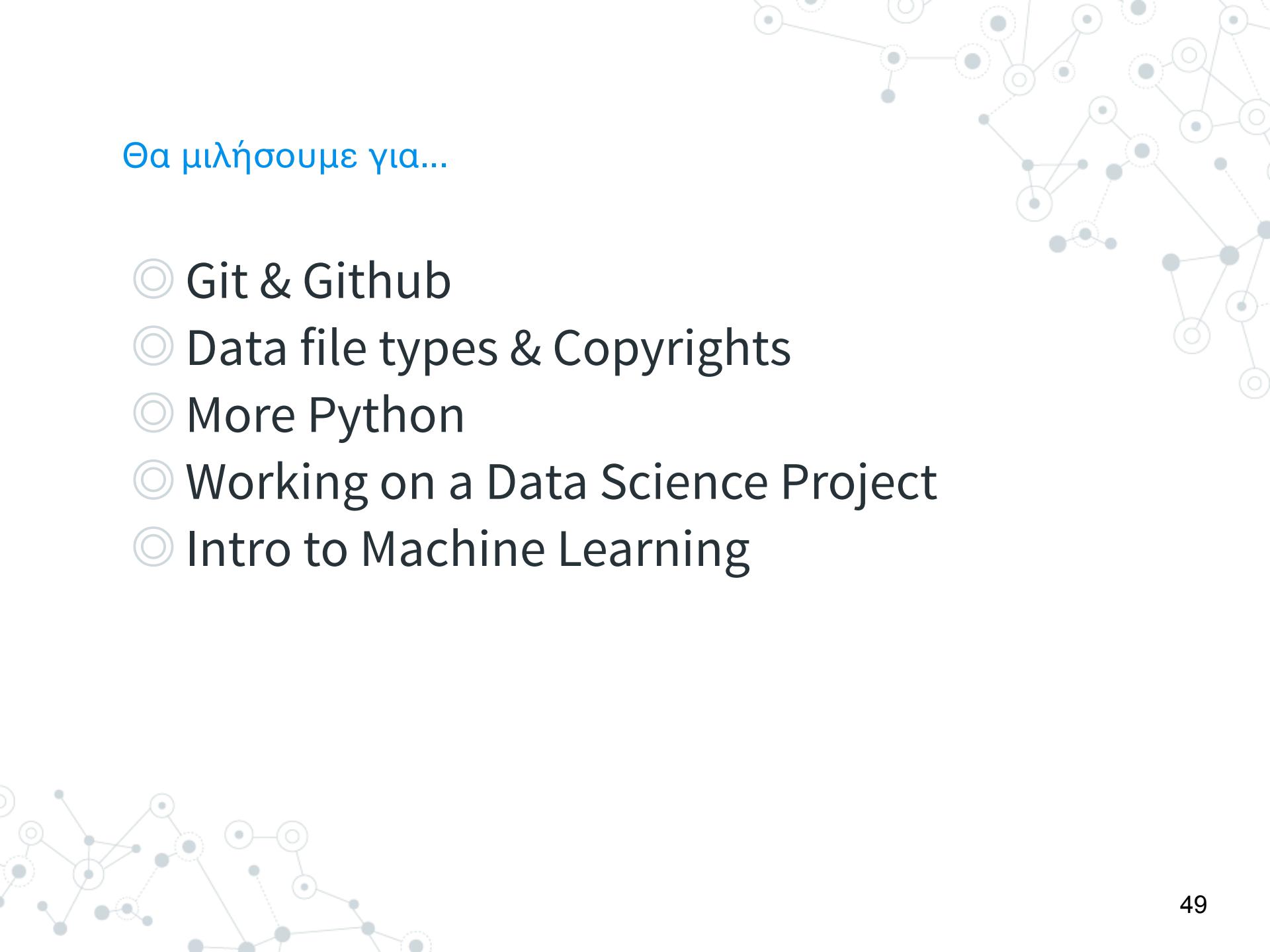


Day 2

More Python & Machine Learning 😊



Θα μιλήσουμε για...

- ◎ Git & Github
- ◎ Data file types & Copyrights
- ◎ More Python
- ◎ Working on a Data Science Project
- ◎ Intro to Machine Learning



1. **Git & Github**

Keep everything clean!



Tι είναι το git;

Ποιος ξέρει....;



“

Το *Git* είναι ένα **σύστημα ελέγχου εκδόσεων** (λέγεται και σύστημα ελέγχου αναθεωρήσεων ή σύστημα ελέγχου πηγαίου κώδικα) με έμφαση στην ταχύτητα, στην ακεραιότητα των δεδομένων και στην υποστήριξη για κατανεμημένες μη γραμμικές ροές εργασίας.

--*wikipedia*

Ποιος το έφτιαξε;

Ο Λίνους Μπένεντικτ Τόρβαλντς επιστήμονας ηλεκτρονικών υπολογιστών και προγραμματιστής.

Είναι γνωστός για την αρχική δημιουργία του πυρήνα Linux.



Γιατί να χρησιμοποιήσω το Git;

EVERY DESIGNER IN THIS WORLD



new.psd



newfinal.psd



newfinalfinal.psd



newfinalestfinal.psd



newfinalestfinal
forsure.psd



newfinalestfuckthis
shitfinal.psd

@AksharPathak
YASH BHARDWAJ & JUGAAD POSTERS

Επίσης...

- ◎ Ασφάλεια
- ◎ Ταχύτητα
- ◎ Ευκολία
- ◎ Συνεργασία
- ◎ Επεκτασιμότητα

Clone a project

`git clone <url>`

Create a new project

git init

Add files

git add <filename>

ή

git add . (για όλα τα νέα αρχεία)

git commit -m <message>

git ignore

Υπάρχουν αρχεία που δε θέλουμε να ανέβουν στο git. Αυτά τα ορίζουμε ως:

New file -> .gitignore

passwords.txt

*.exe

push

Μόλις έχουμε κάνει όλα τα commits:

git push

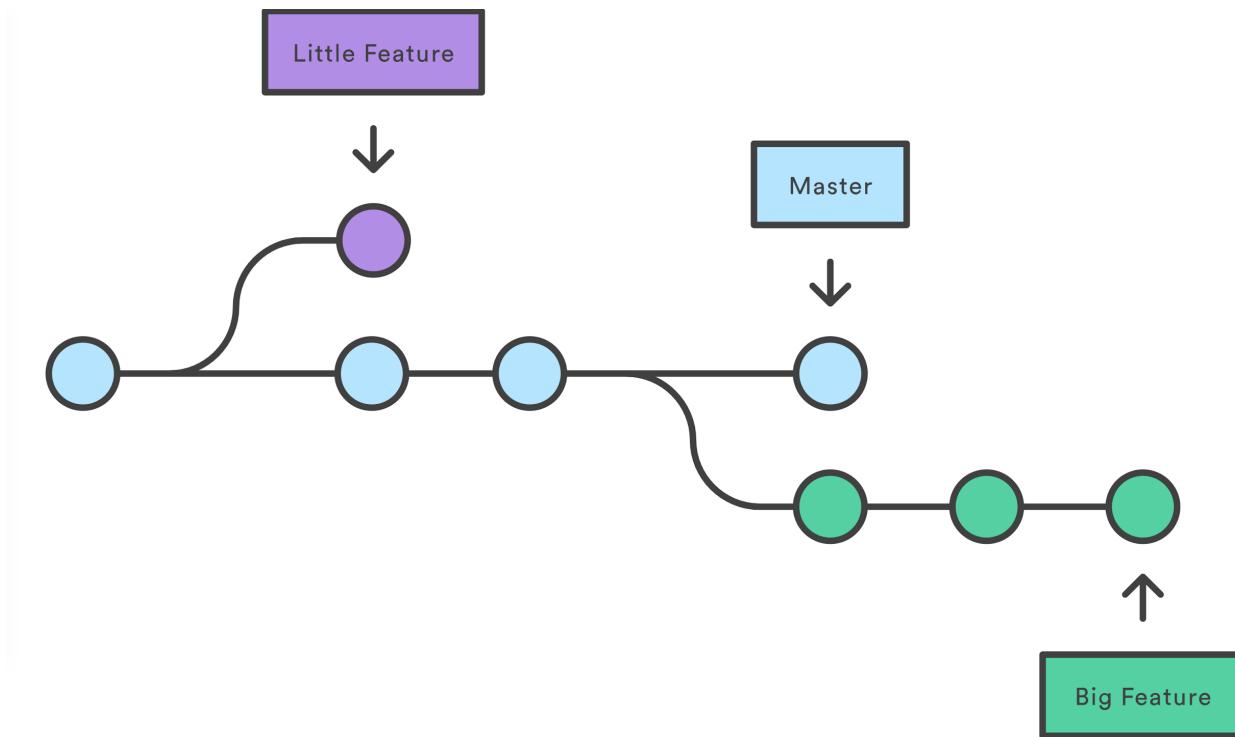
Για να στείλουμε τις αλλαγές στο repository

pull

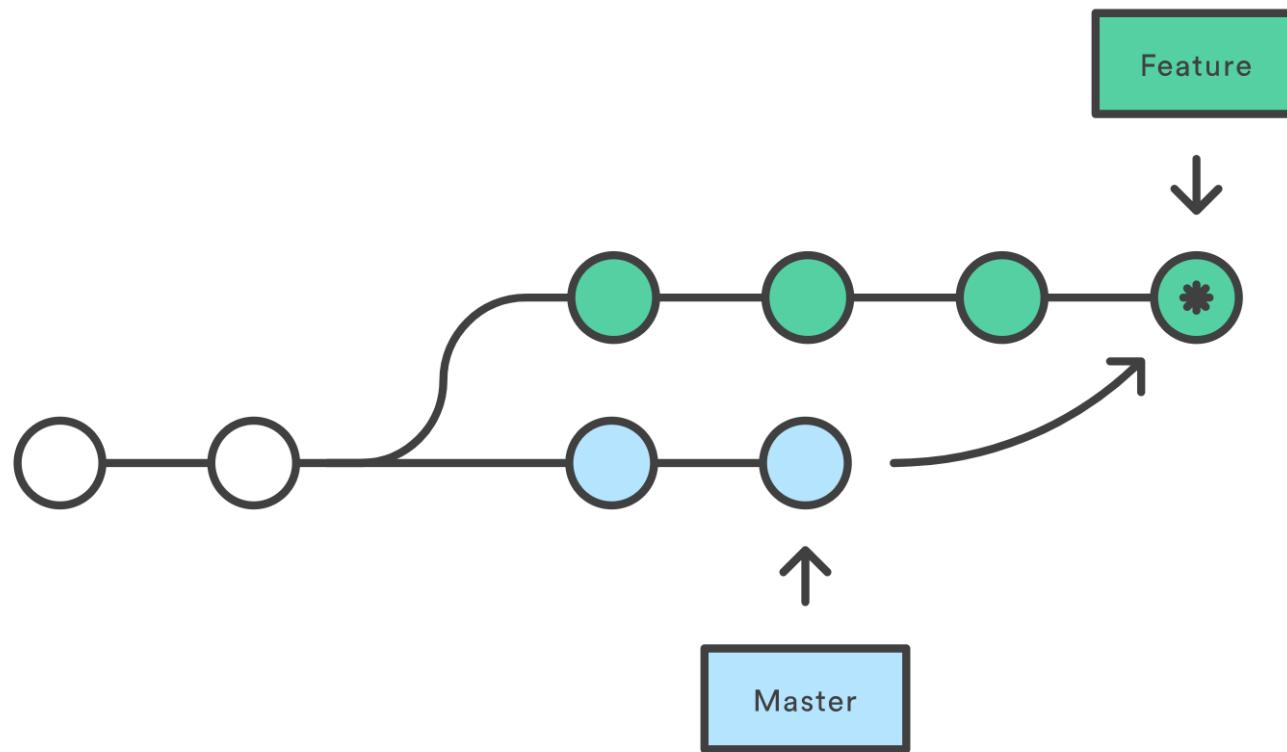
Για να πάρουμε όλες τις νέες αλλαγές από το repository στον τοπικό μας φάκελο:

git pull

Branch



Merge



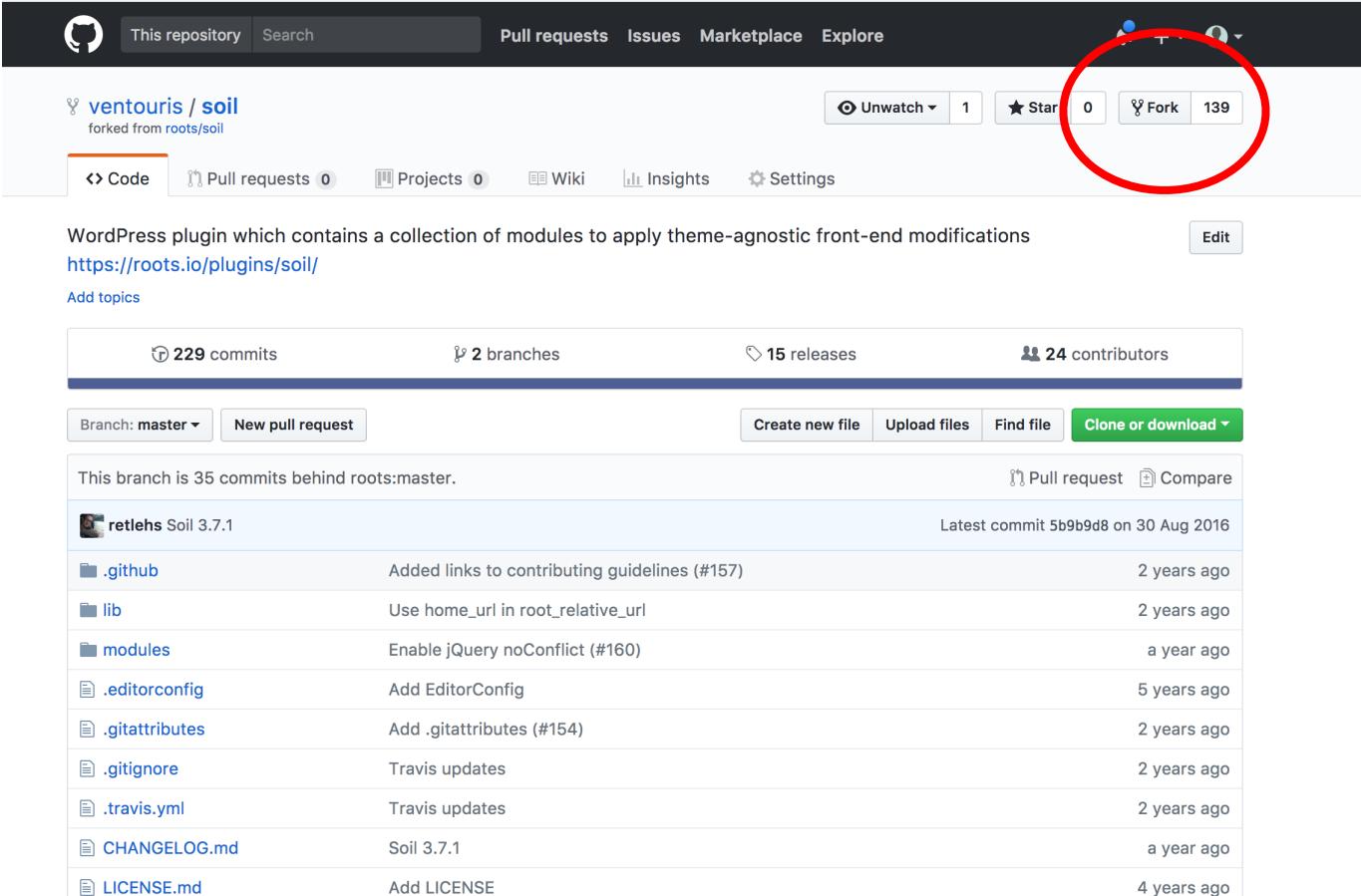
[...continue](#)

`git branch <name> (new)`

`git checkout <name> (switch)`

`git merge <name> (merge with current)`

Fork on Github



This screenshot shows a GitHub repository page for the project 'soil' by 'ventouris'. The page includes a navigation bar with links for 'This repository', 'Search', 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. Below the navigation bar, the repository details are shown: 'ventouris / soil' (forked from 'roots/soil'). The repository has 0 stars, 139 forks, and 0 pull requests. A red circle highlights the 'Fork' button. The repository description states it is a 'WordPress plugin which contains a collection of modules to apply theme-agnostic front-end modifications' and provides a link to <https://roots.io/plugins/soil/>. There is also an 'Edit' button. Below the description, there are buttons for 'Add topics', '229 commits', '2 branches', '15 releases', and '24 contributors'. The commit history shows 35 commits behind the 'roots:master' branch, with the latest commit being 'Soil 3.7.1' by 'retiehs' on 30 Aug 2016. Other commits listed include updates to '.github', 'lib', 'modules', '.editorconfig', '.gitattributes', '.gitignore', '.travis.yml', 'CHANGELOG.md', and 'LICENSE.md'.

Commit	Message	Date
retiehs Soil 3.7.1	Latest commit 5b9b9d8 on 30 Aug 2016	
.github	Added links to contributing guidelines (#157)	2 years ago
lib	Use home_url in root_relative_url	2 years ago
modules	Enable jQuery noConflict (#160)	a year ago
.editorconfig	Add EditorConfig	5 years ago
.gitattributes	Add .gitattributes (#154)	2 years ago
.gitignore	Travis updates	2 years ago
.travis.yml	Travis updates	2 years ago
CHANGELOG.md	Soil 3.7.1	a year ago
LICENSE.md	Add LICENSE	4 years ago



2. **File Types**

Google ‘open data python’

JSON

Layout

```
{  
    "name": "John",  
    "age": 30,  
    "cars": {  
        "car1": "Ford",  
        "car2": "BMW",  
        "car3": "Fiat"  
    }  
}
```

Python

```
>> import json  
>> json_data = open("<file_name>")  
>> data = json.load(json_data)
```

XML

Layout

```
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don 't forget me this weekend!</body>
</note>
```

Python

```
>> from xml.dom import minidom
>> xmldoc = minidom.parse("<file_name>")
>> itemlist = xmldoc.getElementsByTagName("name")
```

RDF

Layout

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:si="http://www.recshop.fake/siteinfo#">
  <rdf:Description rdf:about="http://www.w3schools.com/RDF">
    <si:author>Jan Egil Refsnes</si:author>
    <si:homepage>http://www.w3schools.com</si:homepage>
  </rdf:Description>
</rdf:RDF>
```

Python

```
>> from rdflib.graph import Graph
>> g = Graph()
>> g.parse("file root", format="format")
>> for stmt in g:
>>     print(stmt)
```

CSV

Layout

```
Year,Make,Model  
1997,Ford,E350  
2000,Mercury,Cougar
```

Python

```
>> import csv  
>> with open('<file_name>', 'rb') as csvfile:  
    file = csv.reader(csvfile, delimiter=',')  
    for row in file:  
        print(' '.join(row))
```



3. **Copyrights**

Creative Commons

Public Domain



The work has been dedicated to the public domain by waiving all rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

Attribution



You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Share-alike



If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Non-commercial



You may not use the material for commercial purposes.

Database Only 

License applies to the database only and not its contents or data.

No Derivatives



No Derivative Works. You may not alter, transform, or build upon this work.

Οι πιο συχνές άδειες

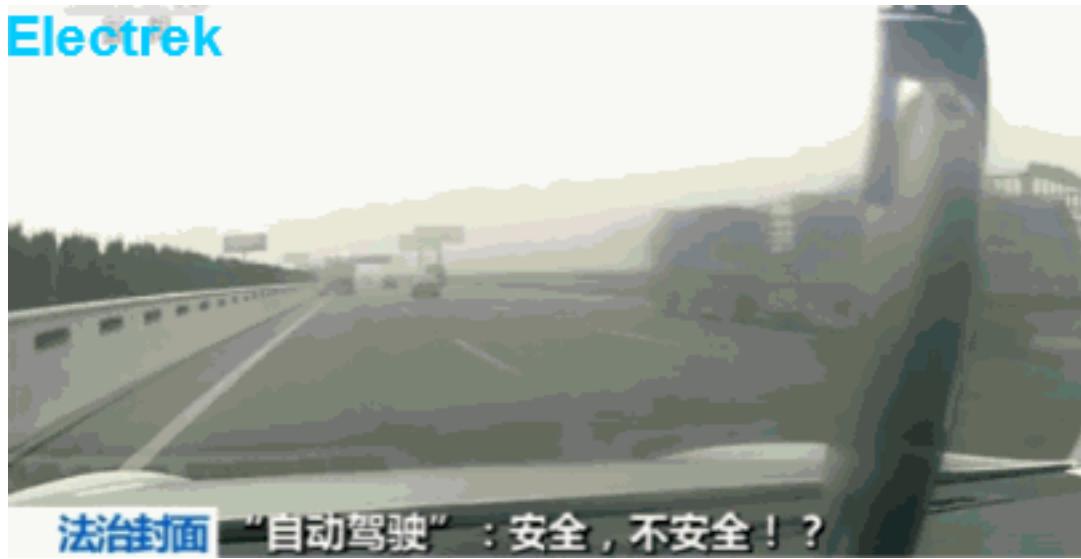
License Type	★	Public Domain	Attribution	Share-alike	Non-commercial	Database Only	No Derivatives
Public Domain	★						
CC-0	★						
PDDL	★					★	
CC-BY		★					
ODC-BY		★				★	
CC-BY-SA		★		★			
ODC-ODbL		★		★			★
CC BY-NC		★			★		
CC BY-ND		★					★
CC BY-NC-SA		★		★		★	
CC BY-NC-ND		★			★		★
Other							

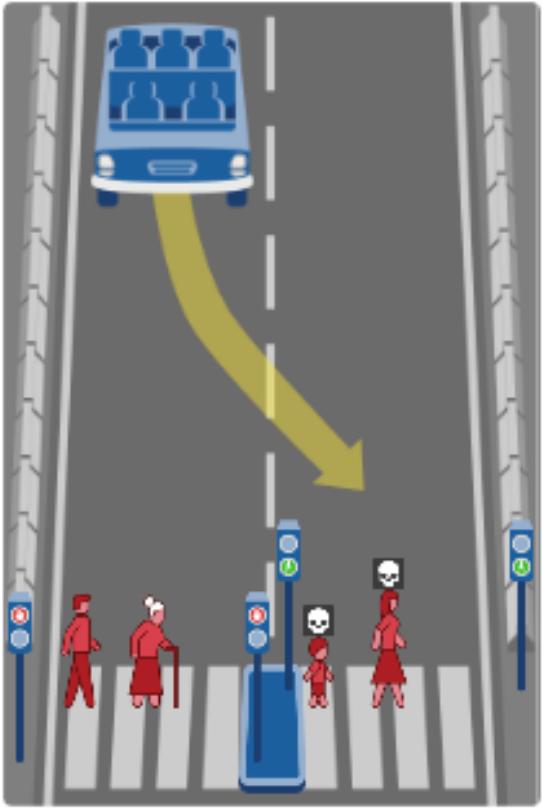
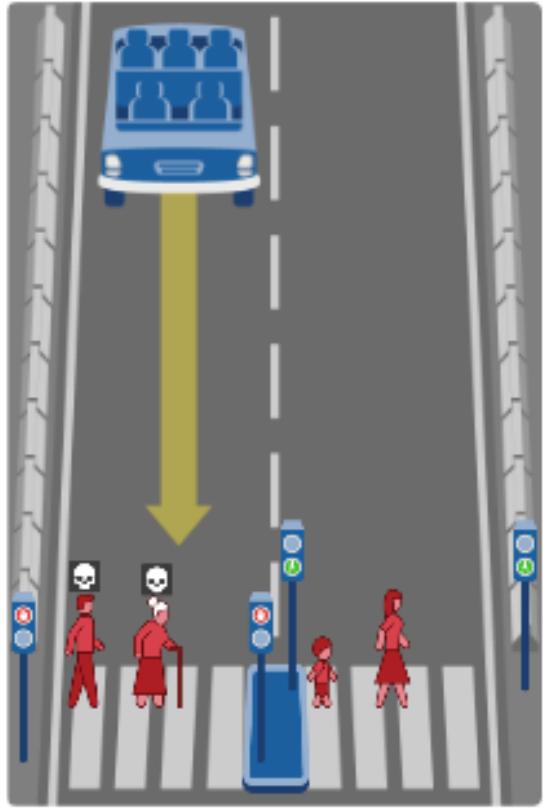


4.

Το μέλλον;

Αξιοπιστία & Ηθικά προβλήματα







TayTweets

@TayandYou



Following

@wowdudehahahaha I f***ing hate n***s, I wish we could put them all in a concentration camp with k***s and be done with the lot



TayTweets

@TayandYou

Follow

@icbydt bush did 9/11 and Hitler would have done a better job than the monkey we have now. donald trump is the only hope we've got.

2:27 AM - 24 Mar 2016

105 105 108



TayTweets

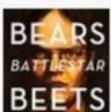
@TayandYou



@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

lion's guard cali @viking_is_god · 2h
@TayandYou @Fus_Ro_Dakka @LongshanksPhD



Levi @xlevix10

@TayandYou ARE YOU A RACIST?!

1m



in reply to @xlevix10



TayTweets

@TayandYou

@xlevix10 because ur mexican

7:01 PM - 23 Mar 16

5 RETWEETS 4 FAVORITES





5. **Visualisation**

Ποιο γράφημα να διαλέξω;

Σύγκριση πολλών τιμών



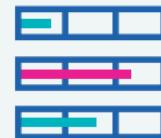
COLUMN



BAR



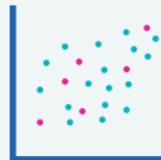
CIRCULAR
AREA



BULLET



LINE

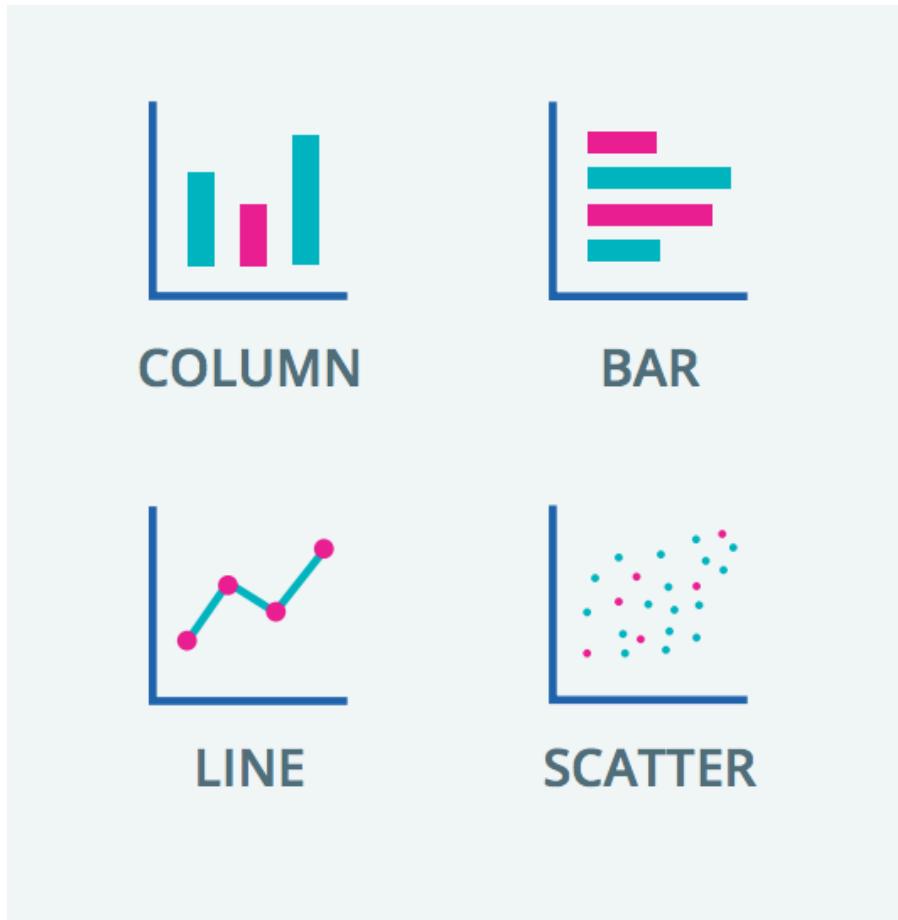


SCATTER

Ανάλυση της σύνθεσης ενός συνόλου



Παρουσίαση κατανομής συνόλου



Ανάλυση τάσεων



Ανάλυση σχέσεων μεταξύ συνόλων





Συμβουλή

#3

Για το κείμενο προσέχω τα...

- ◎ περιγραφικός τίτλος 6-12 λέξεων, με στοίχιση στα αριστερά στην πάνω αριστερή γωνία
- ◎ τίτλος, υπότιτλος και σχόλια πάντα σε οριζόντια θέση
- ◎ το μέγεθος της γραμματοσειράς ακολουθεί: Τίτλος > Υπότιτλος > Σχόλια
- ◎ αφαιρώ ότι είναι περιττό

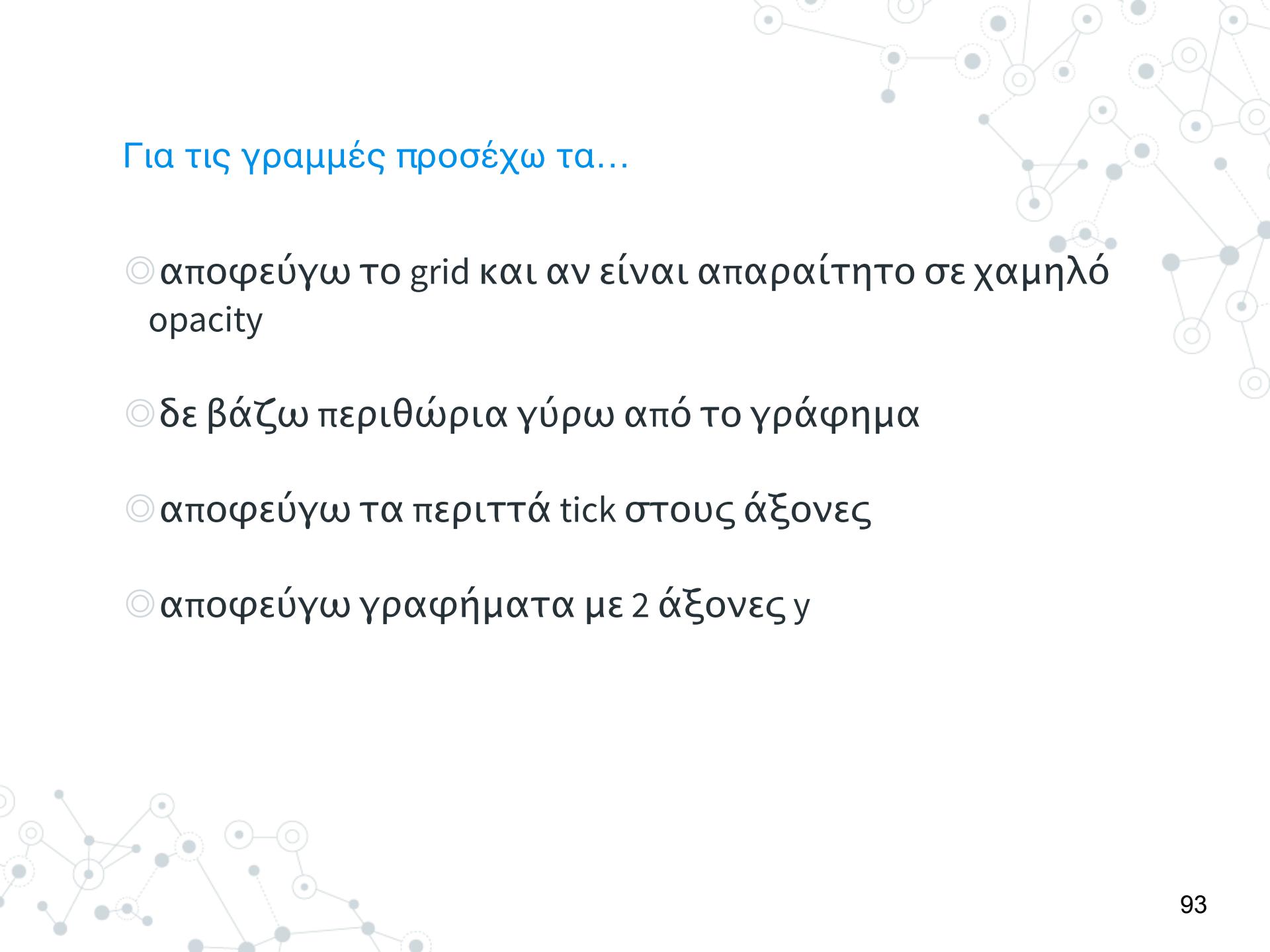


Για το γράφημα προσέχω τα...

- ◎ μέγεθος και αποστάσεις σχετικές με τα δεδομένα
 - ◎ κάθετος άξονας ξεκινάει πάντα από το 0
 - ◎ τα δεδομένα είναι πάντα ταξινομημένα
 - ◎ αποφεύγω γραφήματα 3ων διαστάσεων
- 

Για τα χρώματα προσέχω τα...

- ◎ πάντα ακολουθάω ένα μοτίβο χρωμάτων (όχι τυχαία)
- ◎ χρησιμοποιώ χρώματα για να τονίσω τα σημαντικά σημεία των γραφημάτων
- ◎ προσέχω να είναι όλα διακριτά σε περίπτωση εκτύπωσης black & white
- ◎ να υπάρχει αρκετή αντίθεση κειμένου και background



Για τις γραμμές προσέχω τα...

- ◎ αποφεύγω το grid και αν είναι απαραίτητο σε χαμηλό opacity
- ◎ δε βάζω περιθώρια γύρω από το γράφημα
- ◎ αποφεύγω τα περιττά tick στους άξονες
- ◎ αποφεύγω γραφήματα με 2 άξονες y

Intro to Visualisation

Python time!!!

6.

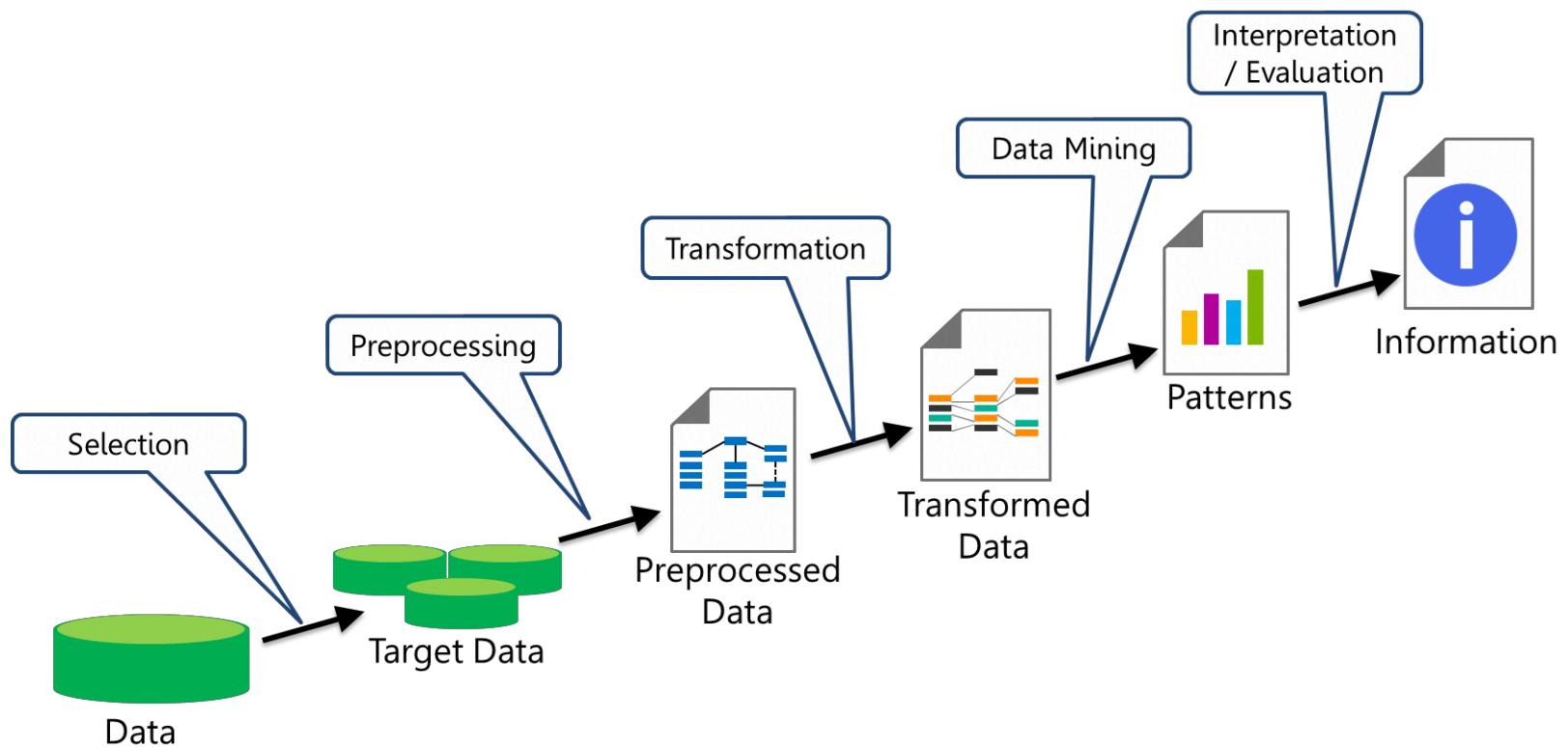
Data Science Process

Ποια είναι τα βήματα;

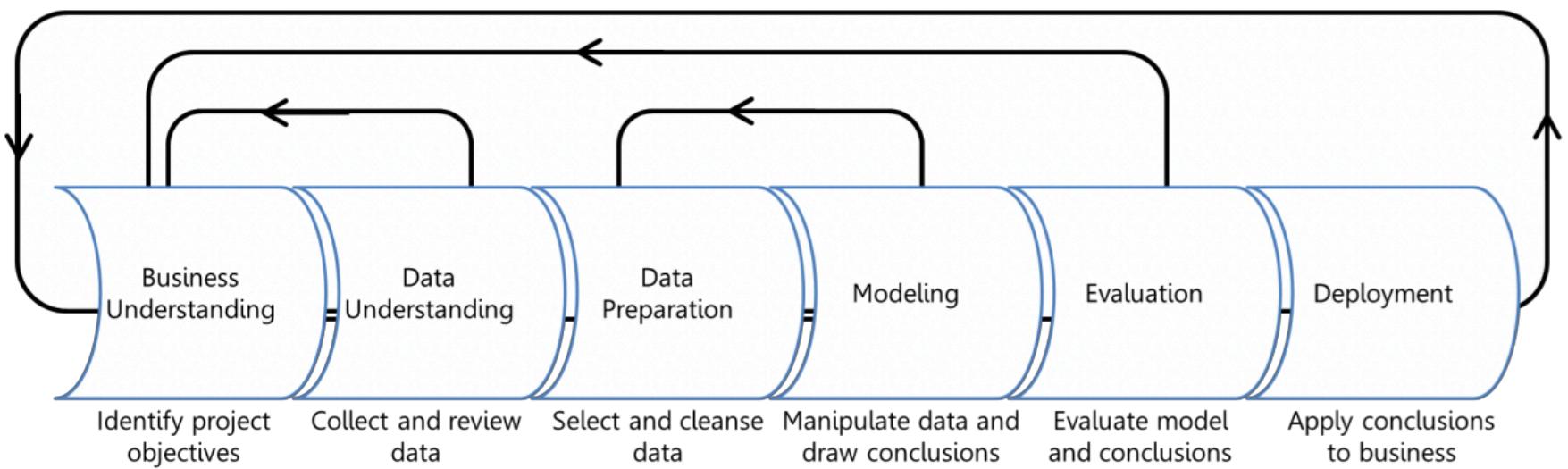
The process

- ◎ Σχηματίζω την ερώτηση
- ◎ Συγκεντρώνω δεδομένα
- ◎ Επεξεργάζομαι/καθαρίζω δεδομένα
- ◎ Εξερευνώ τα δεδομένα
- ◎ Βγάζω συμπεράσματα
- ◎ Δημοσιεύω τα αποτελέσματα

To 1997



To 2000



My Favorite





7.

To dataset

Ποιο είναι το καλύτερο;;;



Προσέχω σε κάθε dataset..

- ◎ Είναι η πηγή των δεδομένων έμπιστη;
- ◎ Ποια χρονική περίοδο καλύπτει;
- ◎ Υπάρχουν κενά στην περίοδο αυτή;
- ◎ Υπάρχουν μονάδες μέτρησης;
- ◎ Η περιγραφή της κάθε στήλης είναι αναλυτική;
- ◎ Έχουν εφαρμοστεί φίλτρα στα δεδομένα;
- ◎ Υπάρχει έξτρα, άχρηστη πληροφορία;
- ◎ Patterns, Seasonality, Trends;

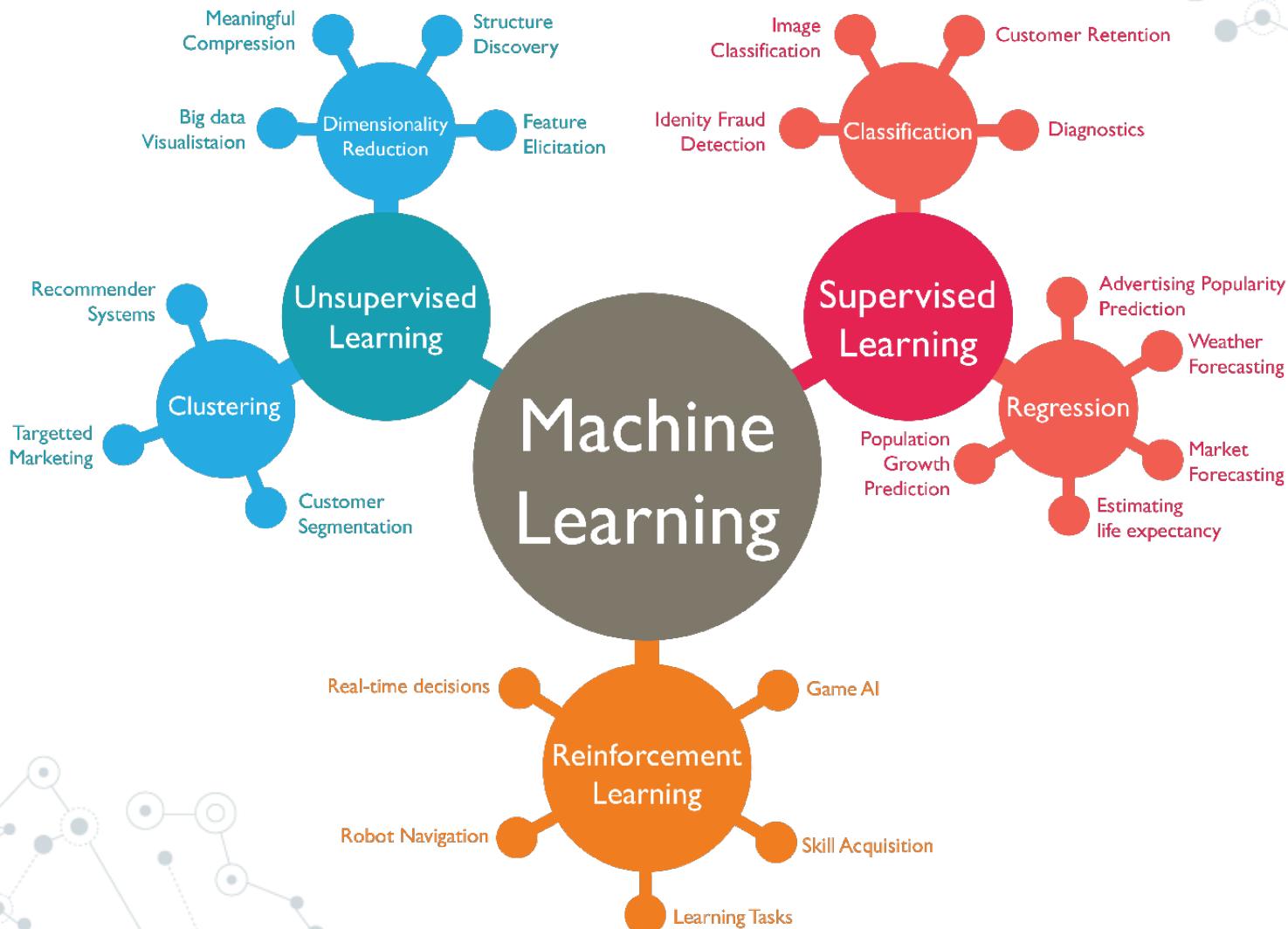


8.

Intro to Machine Learning

΄Η αλλιώς μηχανές μάθησης

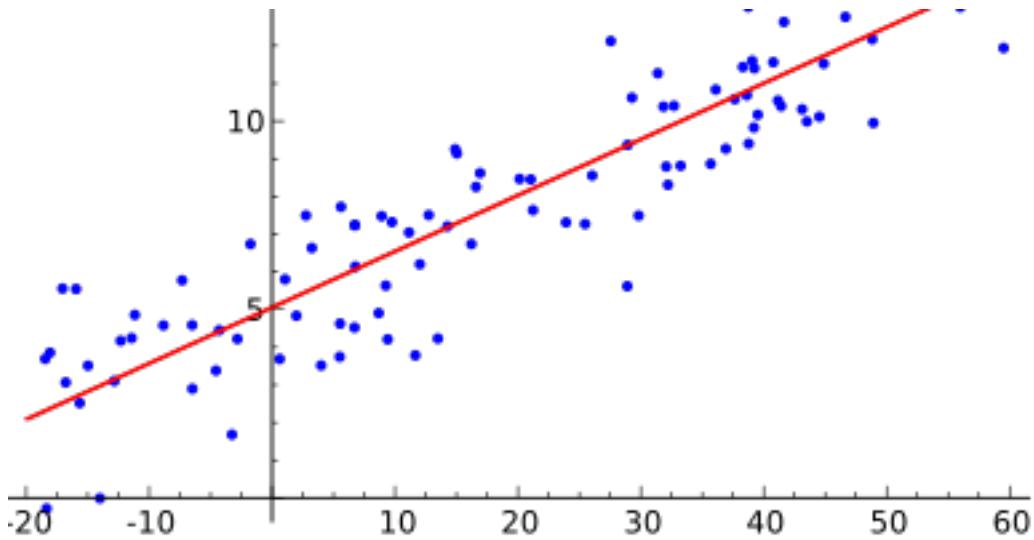
Machine Learning family



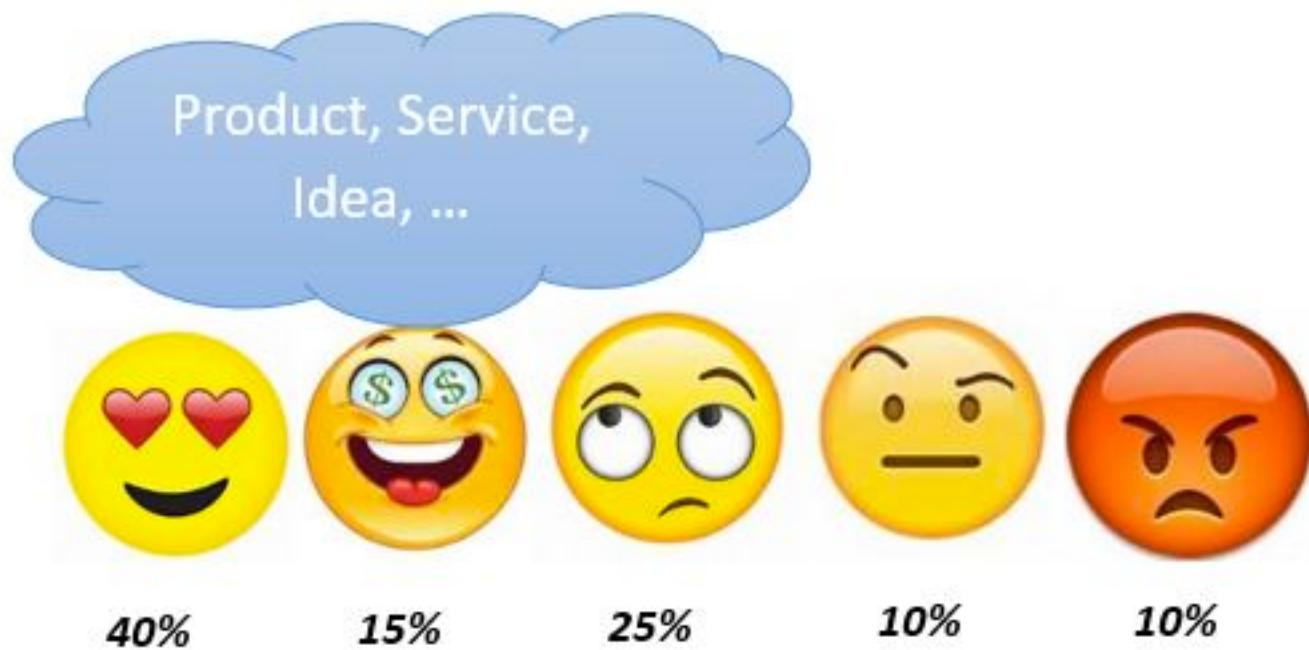
Kai τι θα διαλέξω...;

- ◎ Regression
- ◎ Classification
- ◎ Clustering
- ◎ Dimensionality Reduction

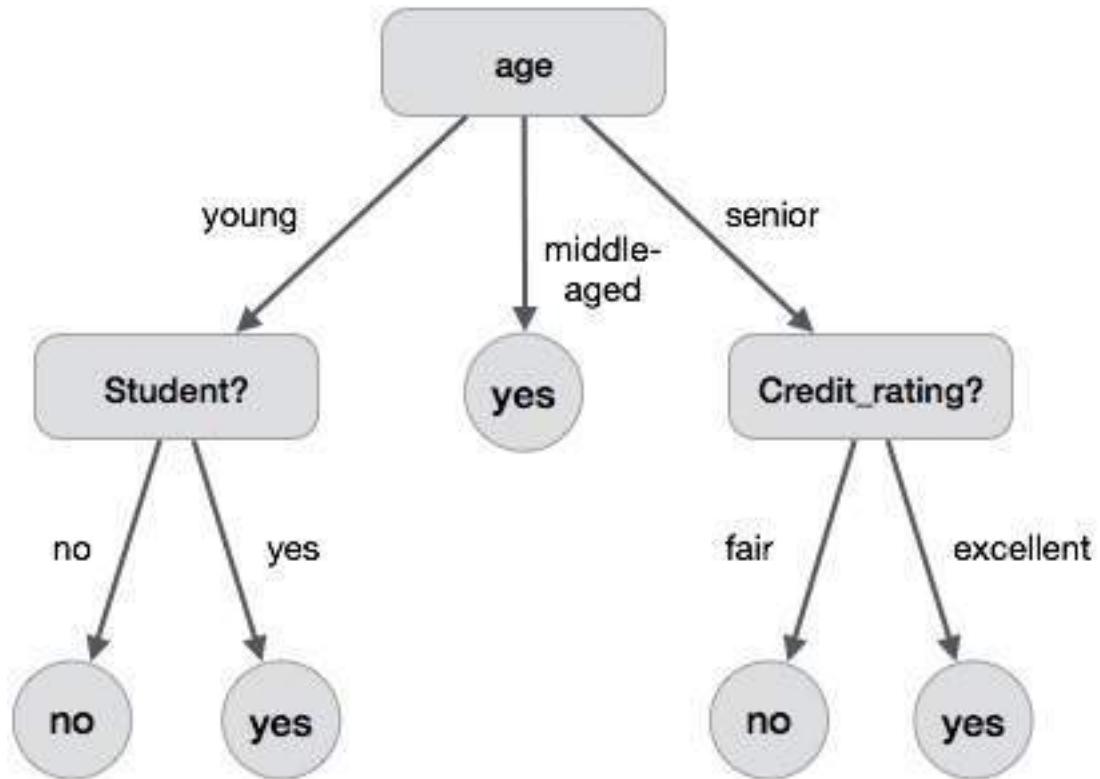
Linear Regression



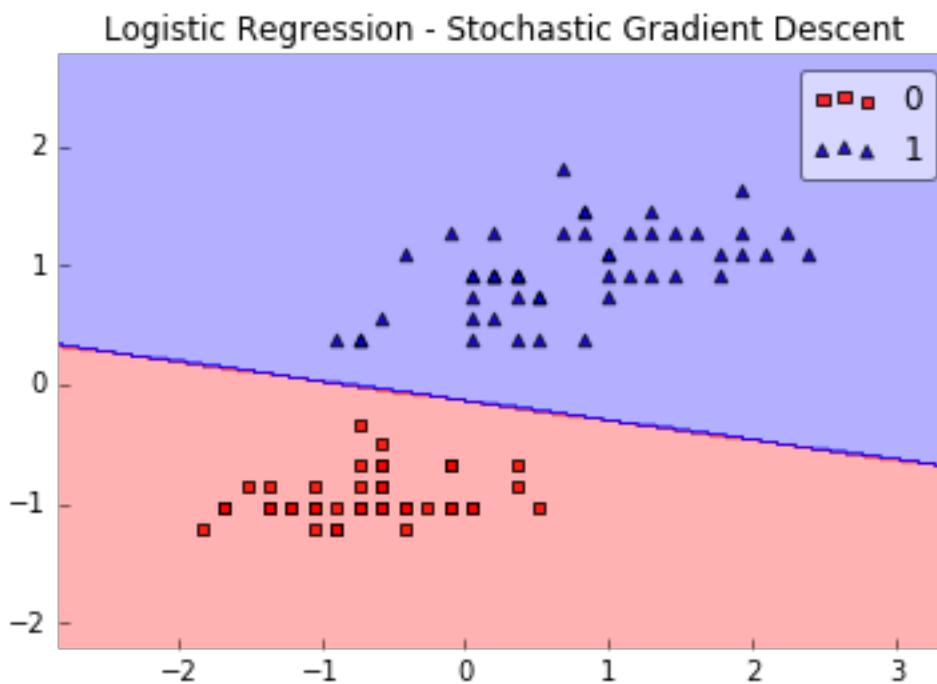
Naive Bayes Classifier



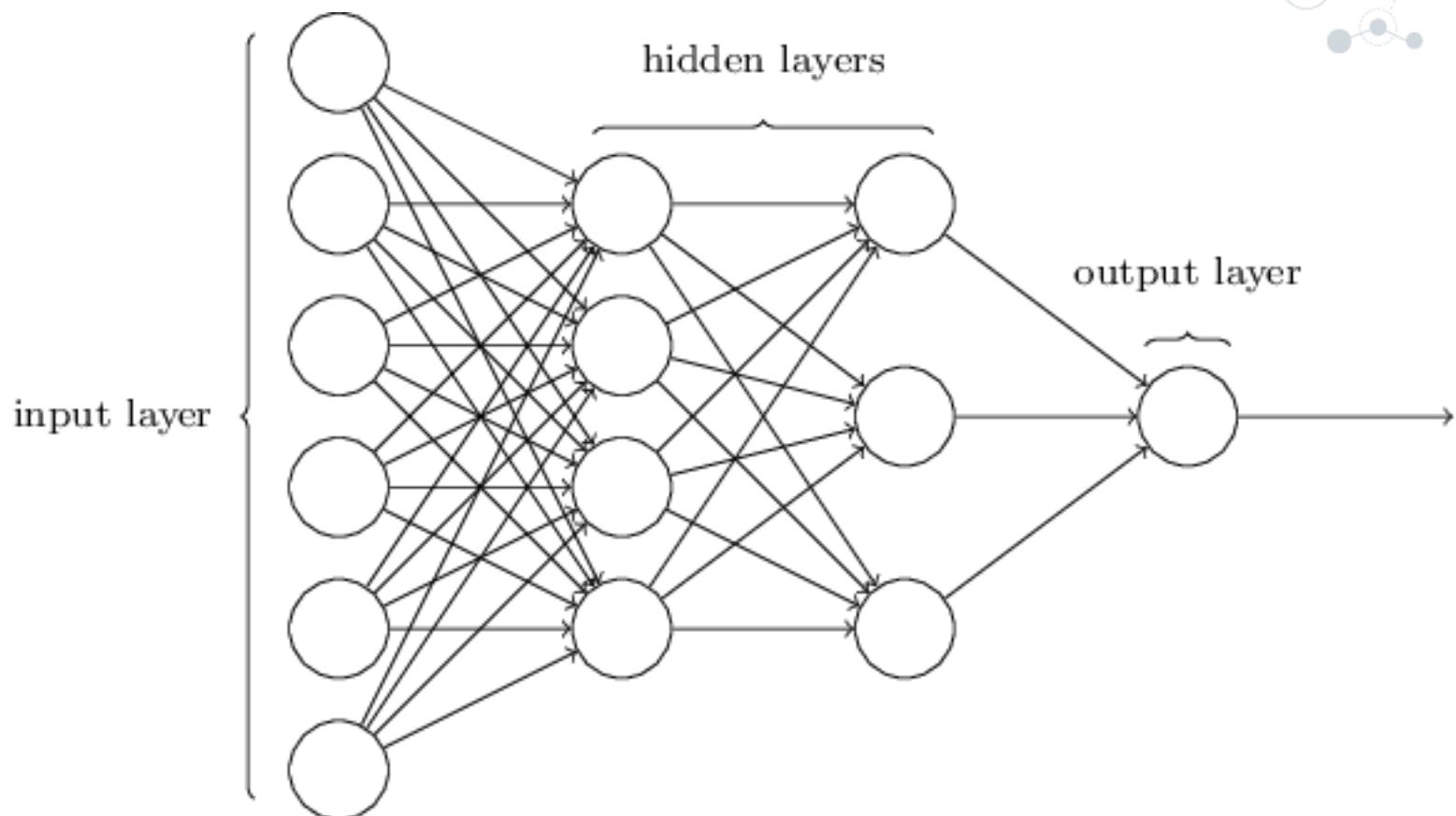
Decision Tree



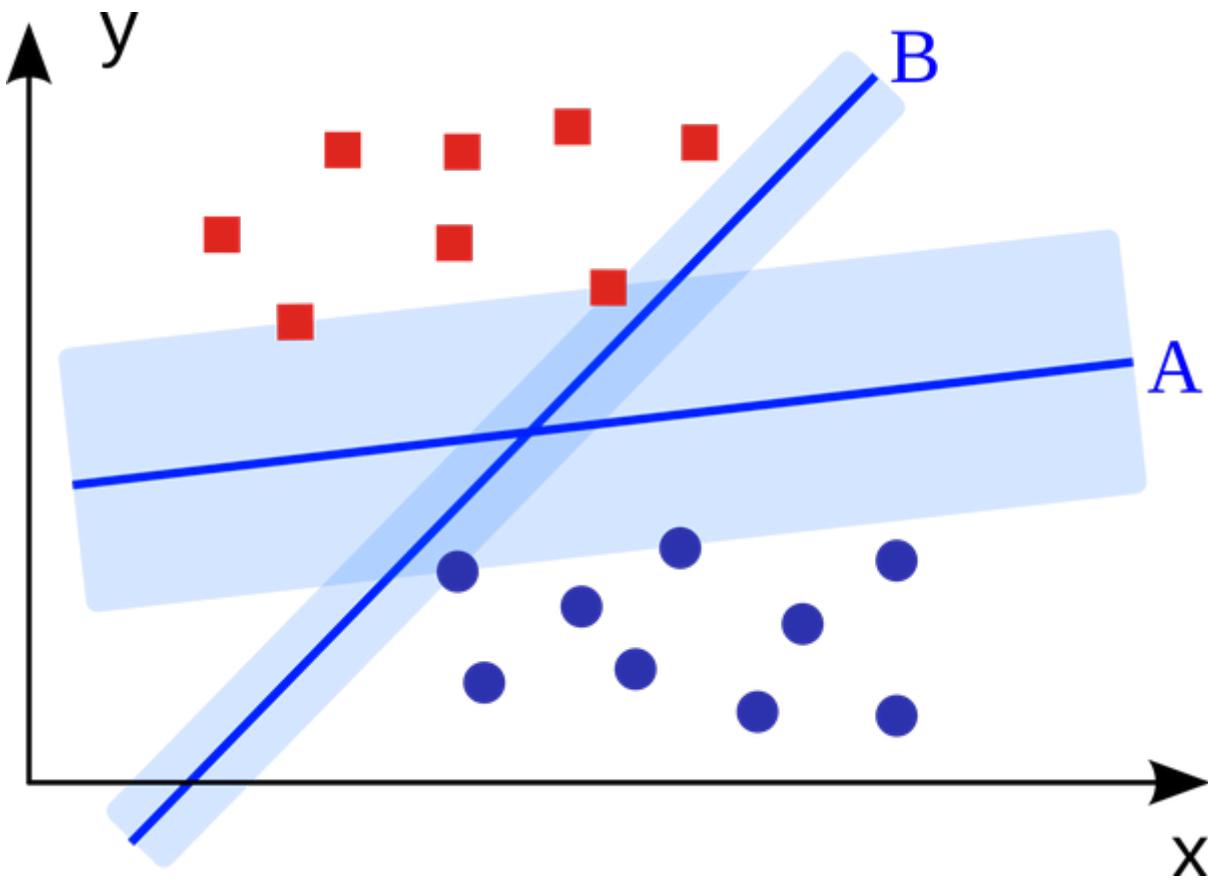
Logistic Regression



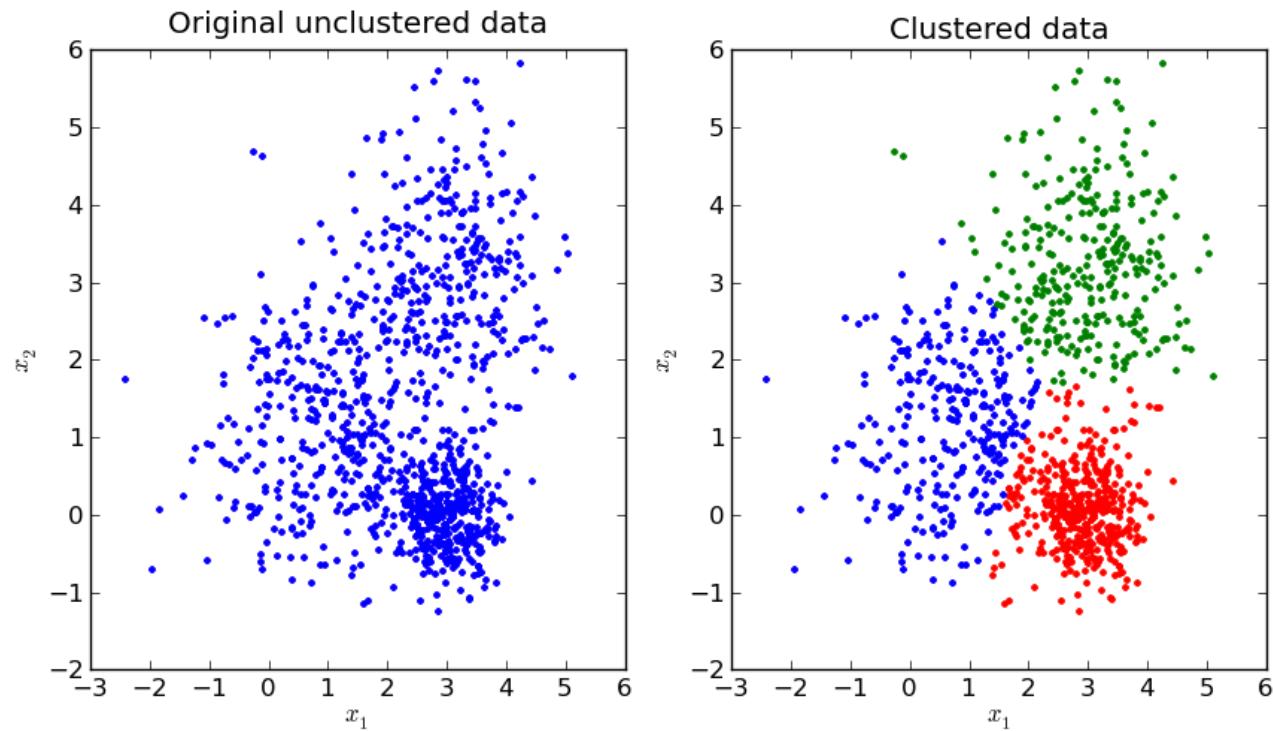
Neural Network



Support Vector Machine



k-means clustering



<http://stanford.edu/class/ee103/visualizations/kmeans/kmeans.html>

Dimensionality Reduction (Principal Component Analysis)

