

CBD2204: Week 2

Takis Zourntos PhD PEng

key roles for the Big Data ecosystem

- Deep Analytical Talent
 - Data Scientists, Machine Learning Engineers
 - employment gap: ~ **200 000** jobs (US ~2020)
- Data Saavy Professionals
 - less technical than data scientists, but can define relevant questions
 - advanced analytics
 - employment gap: ~ **1 500 000** jobs (US ~2020)
- Technology and Data Enablers
 - technical expertise to support analytical projects
 - e.g., provisioning and administering **analytical sandboxes**

what do data scientists do?

1. reframe business challenges as analytics challenges
 - a. what kinds of analytics tools can be brought to bear on the business?
 - b. formulating relevant analytics problems and solving them
2. design and implement **statistical models** and **data mining** techniques on Big Data
 - a. we are going to develop skills using the programming language, R
3. develop insights that lead to actionable recommendations
 - a. thus bridging the gap between theory and practice

characteristics of a data scientist

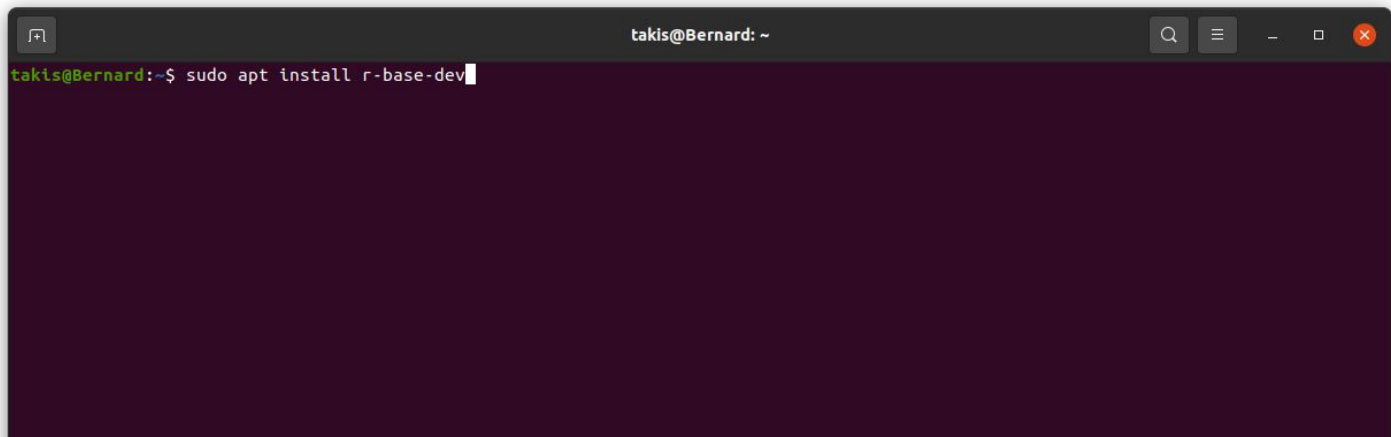


Example: Big Data Analytics

US retailer: Target

- after analyzing consumer purchasing behaviour, Target's statisticians determined that a large portion of revenue came from three different life situations:
 - marriage: when people tend to buy many new products
 - divorce: when people buy new products and change spending habits
 - pregnancy: when people have to buy many new things quickly
- after some analysis, it was determined that pregnancy was the most lucrative of the three
- using data from shoppers, Target was able to identify and predict which of its customers were pregnant, and, in one case, *even before her family knew*
- Target was also able to determine the month of pregnancy
- this allowed them to optimize their inventory, based on the phase of pregnancy (and customer needs)!

an introduction to R



A terminal window titled "takis@Bernard: ~" with standard window controls (search, menu, zoom, close). The terminal has a dark purple background. The prompt "takis@Bernard:~\$" is shown in green, followed by the command "sudo apt install r-base-dev" in white, with a white cursor at the end.

```
takis@Bernard:~$ sudo apt install r-base-dev
```

```
takis@Bernard:~/Downloads/Rstudio$ ls -l
total 117480
-r----- 1 takis takis 120298956 Feb  7 18:02 rstudio-1.4.1103-amd64.deb
takis@Bernard:~/Downloads/Rstudio$ sudo dpkg -i rstudio-1.4.1103-amd64.deb
[sudo] password for takis:
Selecting previously unselected package rstudio.
(Reading database ... 322528 files and directories currently installed.)
Preparing to unpack rstudio-1.4.1103-amd64.deb ...
Unpacking rstudio (1.4.1103) ...
dpkg: dependency problems prevent configuration of rstudio:
 rstudio depends on libclang-dev; however:
  Package libclang-dev is not installed.

dpkg: error processing package rstudio (--install):
 dependency problems - leaving unconfigured
Processing triggers for gnome-menus (3.36.0-1ubuntu1) ...
Processing triggers for desktop-file-utils (0.24-1ubuntu3) ...
Processing triggers for mime-support (3.64ubuntu1) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for shared-mime-info (1.15-1) ...
Errors were encountered while processing:
 rstudio
takis@Bernard:~/Downloads/Rstudio$
```



```
dependency problems - leaving unconfigured
Processing triggers for gnome-menus (3.36.0-1ubuntu1) ...
Processing triggers for desktop-file-utils (0.24-1ubuntu3) ...
Processing triggers for mime-support (3.64ubuntu1) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for shared-mime-info (1.15-1) ...
Errors were encountered while processing:
 rstudio
takis@Bernard:~/Downloads/Rstudio$ sudo apt install libclang-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
You might want to run 'apt --fix-broken install' to correct these.
The following packages have unmet dependencies:
 libclang-dev : Depends: libclang-10-dev (>= 10~) but it is not going to be installed
E: Unmet dependencies. Try 'apt --fix-broken install' with no packages (or specify a solution).
takis@Bernard:~/Downloads/Rstudio$ sudo apt --fix-broken install
Reading package lists... Done
Building dependency tree
Reading state information... Done
Correcting dependencies... Done
The following packages were automatically installed and are no longer required:
  adobe-flash-properties-gtk libllvm10:i386
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  libclang-10-dev libclang-dev
The following NEW packages will be installed:
  libclang-10-dev libclang-dev
0 upgraded, 2 newly installed, 0 to remove and 41 not upgraded.
1 not fully installed or removed.
Need to get 19.4 MB of archives.
After this operation, 185 MB of additional disk space will be used.
Do you want to continue? [Y/n] Y
Get:1 http://ca.archive.ubuntu.com/ubuntu focal/universe amd64 libclang-10-dev amd64 1:10.0.0-4ubuntu1 [19.4 MB]
53% [1 libclang-10-dev 12.8 MB/19.4 MB 66%]
```

```
takis@Bernard:~/Downloads/Rstudio$ sudo dpkg -i rstudio-1.4.1103-amd64.deb
(Reading database ... 325344 files and directories currently installed.)
Preparing to unpack rstudio-1.4.1103-amd64.deb ...
Unpacking rstudio (1.4.1103) over (1.4.1103) ...
Setting up rstudio (1.4.1103) ...
Processing triggers for gnome-menus (3.36.0-1ubuntu1) ...
Processing triggers for desktop-file-utils (0.24-1ubuntu3) ...
Processing triggers for mime-support (3.64ubuntu1) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for shared-mime-info (1.15-1) ...
takis@Bernard:~/Downloads/Rstudio$
```

