Extend your toolbox with R

Szymon Skorupinski
IT-DB
R is...

- Open source version of S ;)
- Programming language
- Computing environment
- Designed especially for data
  - analysis
  - manipulation
  - visualization
Example data set

- From UBS Prices & Earnings 2015 report
- Working hours required to buy iPhone 6 for 70 cities

<table>
<thead>
<tr>
<th>Year</th>
<th>City</th>
<th>Name</th>
<th>What</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Amsterdam</td>
<td>Working time required to buy.</td>
<td>1 iPhone 6 16 GB, in hrs.</td>
<td>49,75</td>
</tr>
<tr>
<td>2015</td>
<td>Athens</td>
<td>Working time required to buy.</td>
<td>1 iPhone 6 16 GB, in hrs.</td>
<td>98,21</td>
</tr>
<tr>
<td>2015</td>
<td>Auckland</td>
<td>Working time required to buy.</td>
<td>1 iPhone 6 16 GB, in hrs.</td>
<td>44,62</td>
</tr>
<tr>
<td>2015</td>
<td>Bangkok</td>
<td>Working time required to buy.</td>
<td>1 iPhone 6 16 GB, in hrs.</td>
<td>149,57</td>
</tr>
<tr>
<td>2015</td>
<td>Barcelona</td>
<td>Working time required to buy.</td>
<td>1 iPhone 6 16 GB, in hrs.</td>
<td>59,06</td>
</tr>
<tr>
<td>2015</td>
<td>Beijing</td>
<td>Working time required to buy.</td>
<td>1 iPhone 6 16 GB, in hrs.</td>
<td>217,8</td>
</tr>
</tbody>
</table>

(…)


Example data set

• From UBS Prices & Earnings 2015 report
• Working hours required to buy iPhone 6 for 70 cities

Year;City;Name;What;Hours
2015;Amsterdam;Working time required to buy..;1 iPhone 6 16 GB, in hrs.;49,75
2015;Athens;Working time required to buy..;1 iPhone 6 16 GB, in hrs.;98,21
2015;Auckland;Working time required to buy..;1 iPhone 6 16 GB, in hrs.;44,62
2015;Bangkok;Working time required to buy..;1 iPhone 6 16 GB, in hrs.;149,57
2015;Barcelona;Working time required to buy..;1 iPhone 6 16 GB, in hrs.;59,06
2015;Beijing;Working time required to buy..;1 iPhone 6 16 GB, in hrs.;217,8
(...
Reading data

data <- read.csv2("data.csv")
hours <- data$Hours
names(hours) <- data$City
Your first bar chart!

barplot(hours)
Better sort data...

hours <- sort(hours)
...and plot it again...

barplot(hours)
...with better x-axis labels...

\texttt{barplot(hours, las = 2)}
... and horizontally...

```r
barplot(
  hours,
  las = 2,
  horiz = TRUE)
```
...plus with colors...

```r
barplot(
  hours,
  las = 2,
  horiz = TRUE,
  col = rainbow(length(hours)))
```
...and other minor changes

barplot(
    hours,
    las = 2,
    horiz = TRUE,
    col = rainbow(length(hours)),
    axis.lty = 1,
    main = "Working time to buy iPhone 6 - in hours")
Visualize median

```
abline(
  v = median(hours),
  lwd = 3)
```
Another chart type

```python
hist(
    hours,
    col = rainbow(7))
```
Or both together?

par(mfrow = c(1,2))

Then draw both graphs again and you get...
Save your work

dev.print(
    pdf,
    "charts.pdf")

• Many other formats available
  • jpg, png, svg, postscript, etc.
More graphics examples

From demo() function
This is only tip of the iceberg

• Of R graphics
  • Quickly create publication ready graphs
    • Only base package shown today
    • Many others available, e.g.: lattice, ggplot2

• Of other vast R capabilities
  • Statistics, analytics, machine learning etc.
  • Extensible by over 7000 packages
    • https://cran.r-project.org/web/packages
Try it!

- R - https://cran.r-project.org
- IDE for R - https://www.rstudio.com
- MOOC, e.g.
  - R Programming on Coursera
- Interactive tutorial about R on R
  - http://swirlstats.com
- Without installing anything – in your browser
  - https://www.datacamp.com/swirl-r-tutorial
Questions?

Thank you!