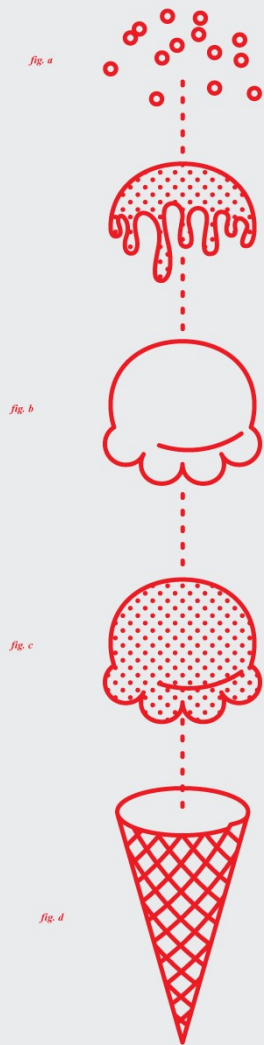*fig. a*

fig. b

fig. c

fig. d

Which programming language to learn - R or Python?

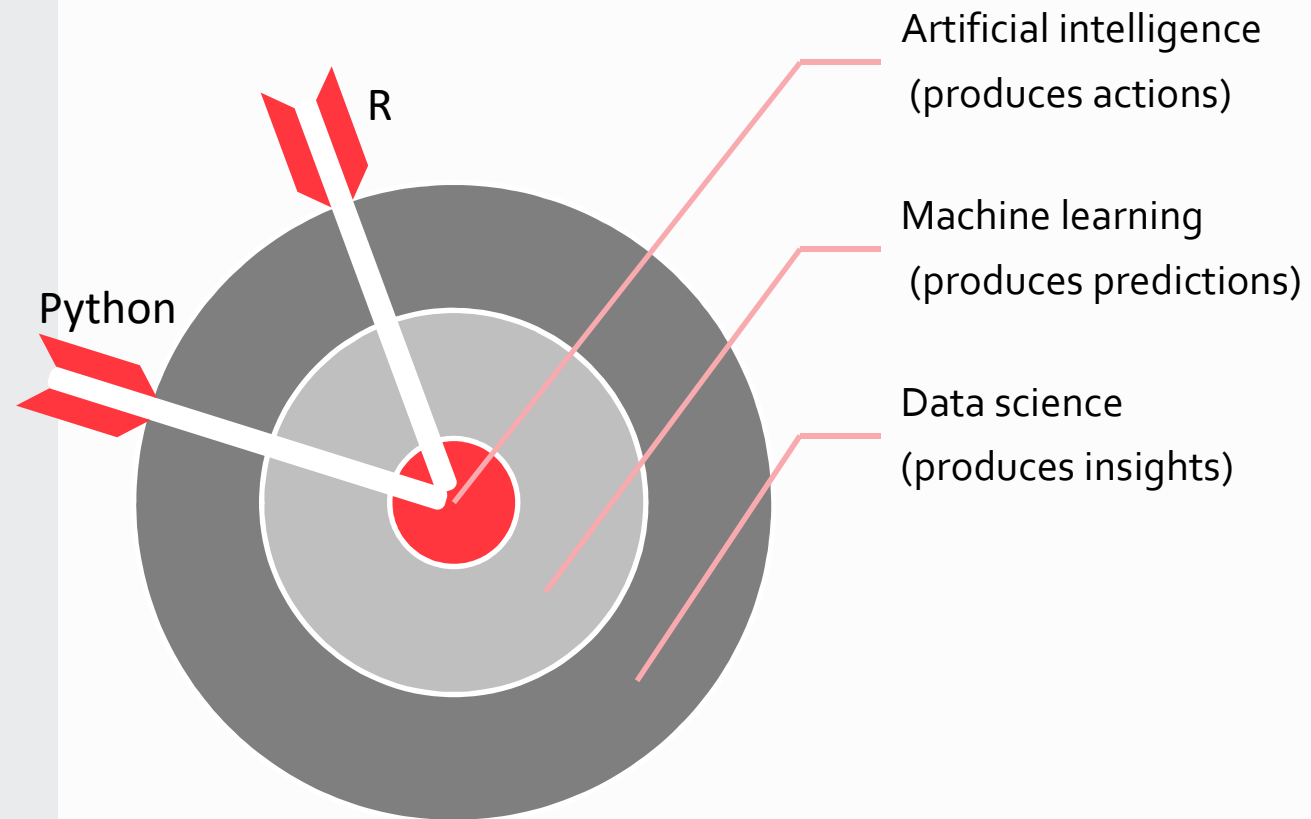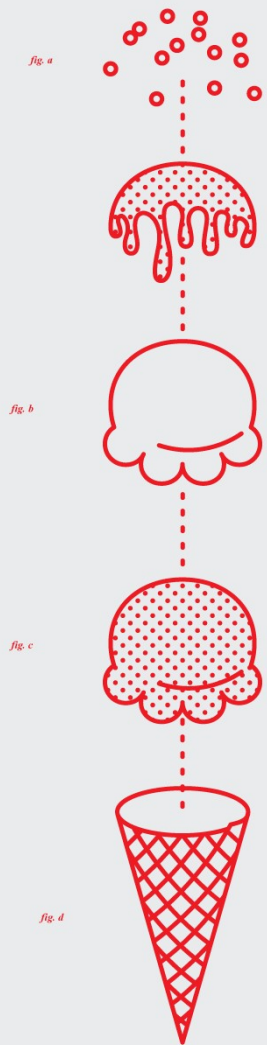Maggie Petrova

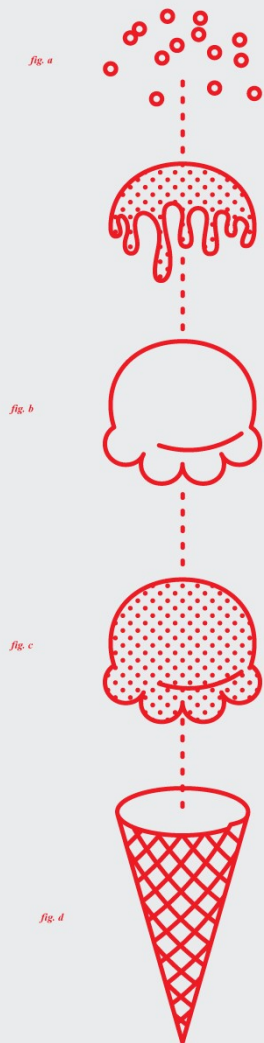Data Scientist @ObjectiveIT

Mar-2018

Outline

- What are the R and Python programming languages used for?

- Which programming language is more popular?

- Pros of R and Python

- And the winner is

- Learning curve of the language

- Where do you start

- My top 3 tips

What are the R and Python programming languages used for?

fig. a

fig. b

fig. c

fig. d

R

Python

Artificial intelligence (produces actions)

Machine learning (produces predictions)

Data science (produces insights)

fig. a

fig. b

fig. c

fig. d

IEEE (Institute of Electrical and Electronics Engineers) Top Programming Languages

| IEEE Spectrum | Trending | Jobs | Open | Custom |

Edit Ranking | Add a Comparison |

Language Types (click to hide)

| 🌐 Web | 📱 Mobile | 🖥 Enterprise | ▦ Embedded |

Language Rank	Types	Spectrum Ranking
1. Python	🌐 🖥	100.0
2. C	📱 🖥 ▦	99.7
3. Java	🌐 📱 🖥	99.4
4. C++	📱 🖥 ▦	97.2
5. C#	🌐 📱 🖥	88.6
6. R	🖥	88.1
7. JavaScript	🌐 📱	85.5
8. PHP	🌐	81.4
9. Go	🌐 🖥	76.1
10. Swift	📱 🖥	75.3

Source: https://spectrum.ieee.org/static/interactive-the-top-programming-languages-2017

Popularity on Stack Overflow
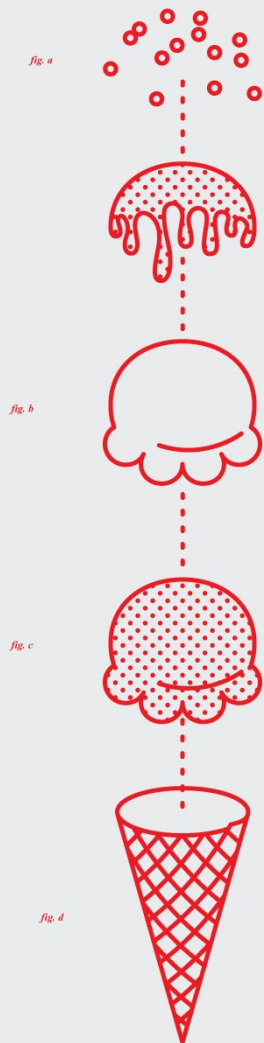
python × 917418

a dynamic and strongly typed programming language designed to emphasize usability. Two similar but mostly incompatible versions

890 asked today, 5429 this week

r × 228523

a free, open-source programming language and software environment for statistical computing, bioinformatics, visualization and

236 asked today, 1440 this week

fig. a

fig. b

fig. c

fig. d

Who is using Python and/or R

Python

R

fig. a

fig. b

fig. c

fig. d

Pros of R and Python

What is R good at?

R is a free, open-source programming language and software environment for statistical computing, visualisation and general computing.

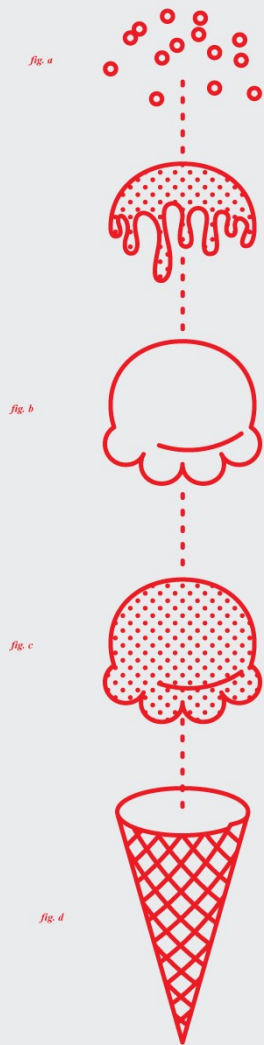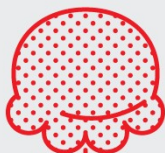R is mainly used when the data analysis task requires standalone computing or analysis on individual servers. It's great for exploratory work, and it's handy for almost any type of data analysis because of the huge number of packages.

What is Python good at?

Python is a dynamic and strongly typed programming language that is used for a wide range of applications. Python is used from testing microchips at Intel, to powering Instagram, to building video games with the PyGame library. It is a general-purpose, high-level programming language that is designed to emphasize usability.

Python allows programmers to express concepts in fewer lines of clear code.

You can use Python when your data analysis tasks need to be integrated with web apps or if statistics code needs to be incorporated into a production database. Being a fully fledged programming language, it's a great tool to implement algorithms for production use.
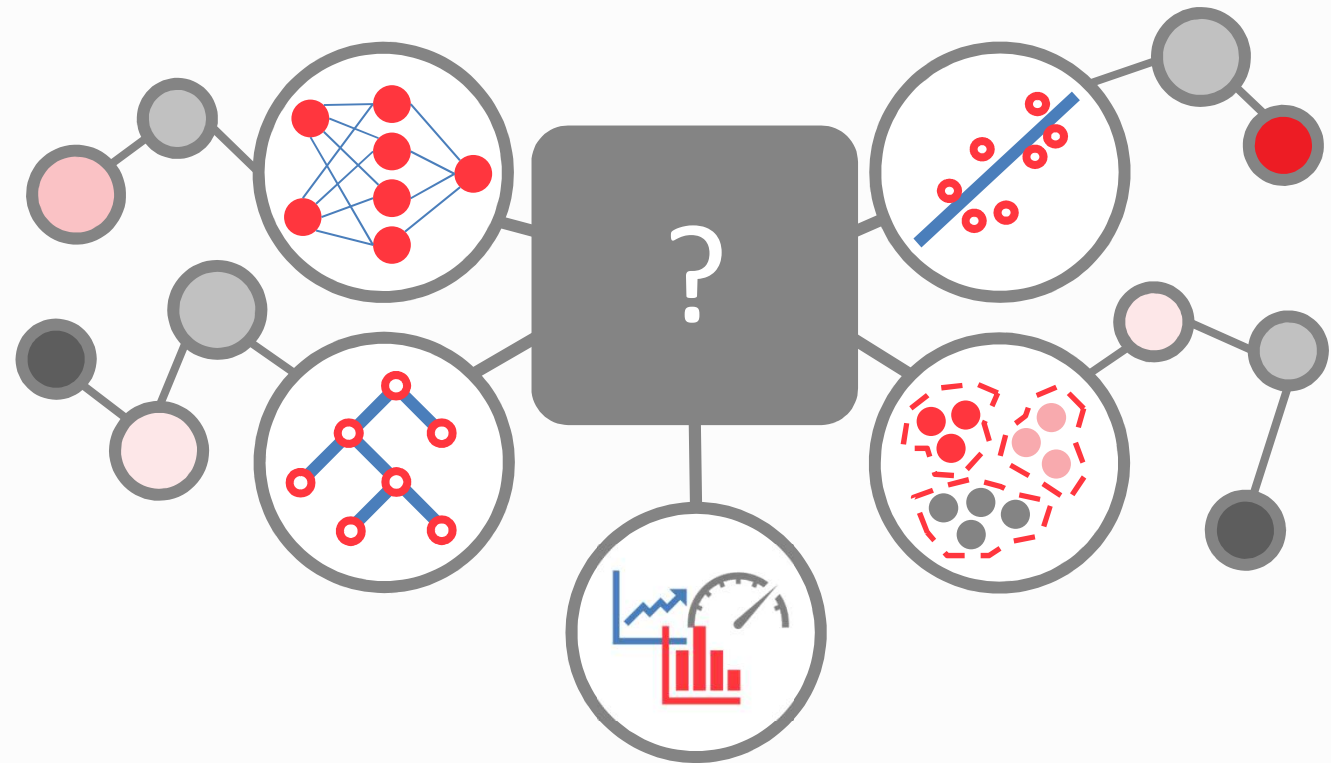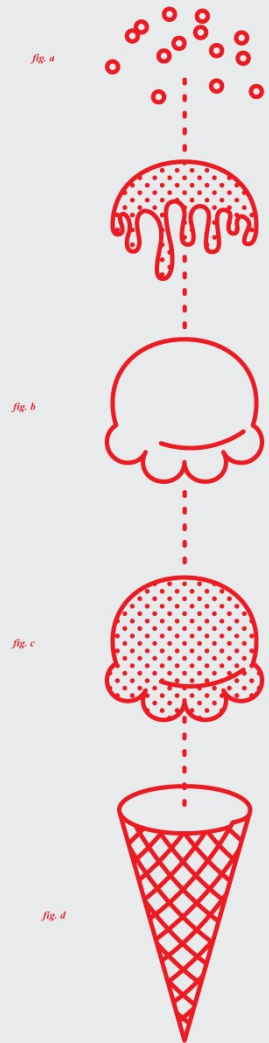
And the winner is?

fig. a

fig. b

fig. c

fig. d

Transition to R or Python
(setting the right expectations)

Your background	R	Python
Statistician		
Developer		
Non-statistician Analyst (I love Excel)		

fig. a

fig. b

fig. c

fig. d

R learning curve

R Learning curve

Python difficulty

Languages ARRANGED BY **Difficulty**

MORE Difficult →

OH, HEY, I DIDN'T SEE YOU GUYS ALL THE WAY OVER THERE.

Basic   Python   C++   Lisp   Haskell   Malbolge
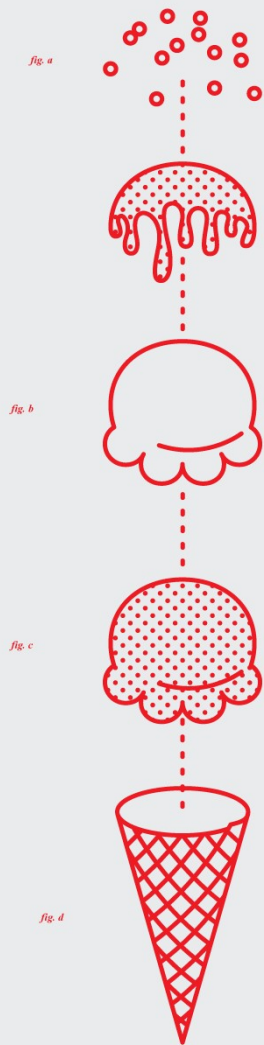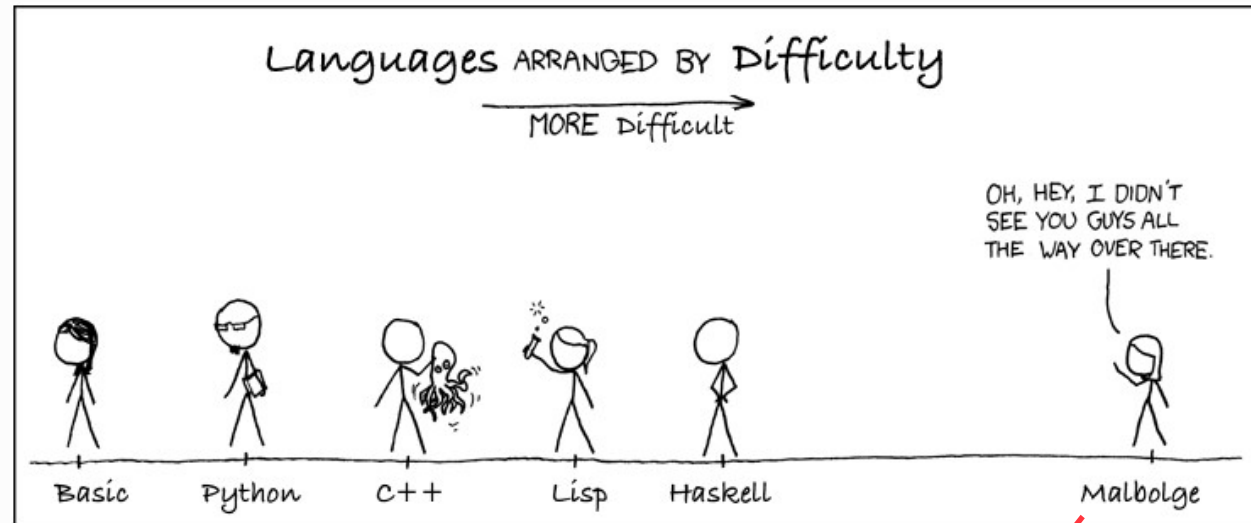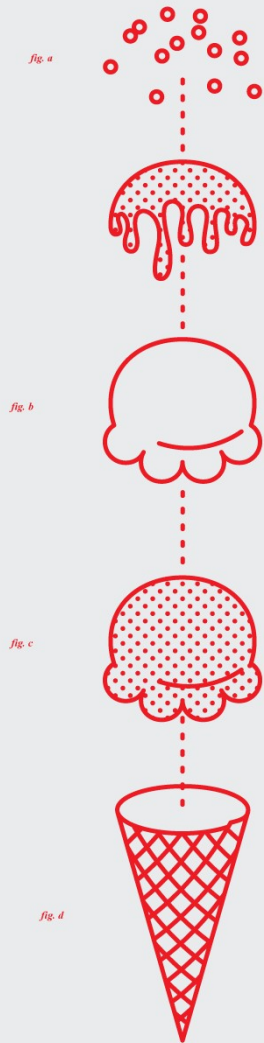
Malbolge is a programming language named after the eighth circle of hell in Dante's Inferno. It was specifically designed to be almost impossible to use, via a counter-intuitive 'crazy operation', and self-altering code.
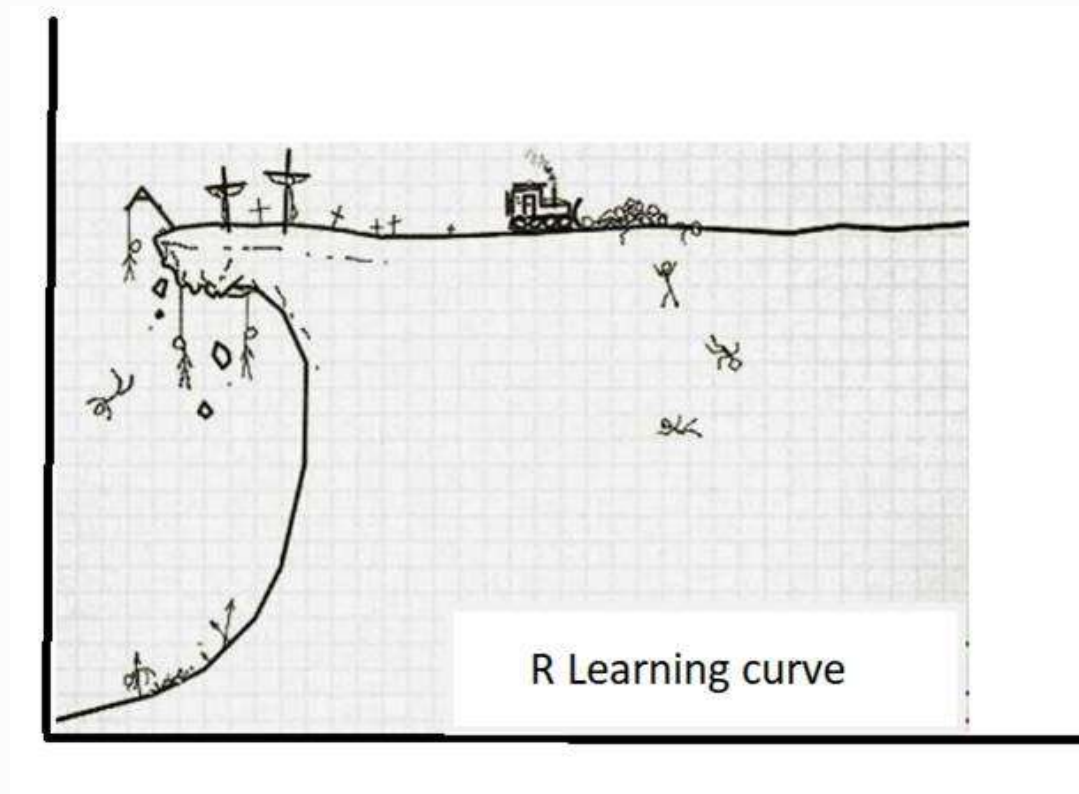
fig. a

fig. b

fig. c

fig. d
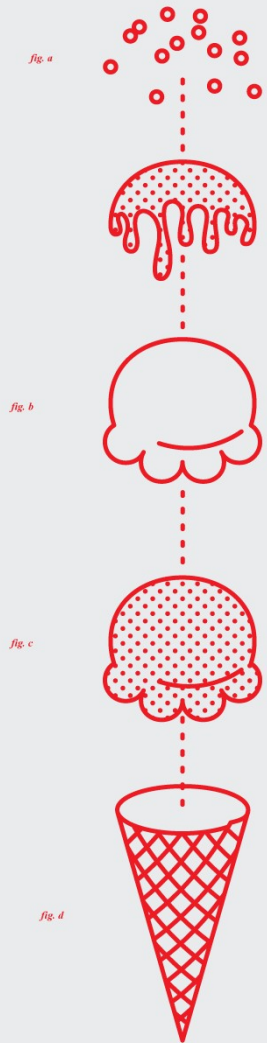
Where do you start?

Jupyter notebook
Spyder
scikit-learn
zoo
library
Python
data.table
packages
stringr
R
caret
pandas
ggplot2
NumPy
IDE
matplotlib
dplyr
RStudio

Where do you start?

| Key libraries	R	Python
Integrated development environment		
IDE	*RStudio*	*Jupyter Notebook, Spyder, PyCharm, etc.*
Popular packages /libraries		
Data manipulation & computing	dplyr, plyr, data.table	Pandas, NumPy, SciPy
Text mining	stringr	string
Time series	zoo, xts	Prophet
Visualisation and reporting results	ggvis, lattice, ggplot2, shiny, RMarkdown	Matplotlib, Seaborn, Plotly
Machine learning	caret, randomForest, nnet	scikit-learn, Keras, TensorFlow, NLTK
Data scraping	rvest	Scrapy

fig. a

fig. b

fig. c
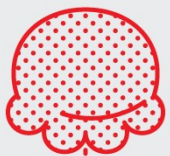
fig. d

Where to learn

fig. a

fig. b

fig. c

fig. d
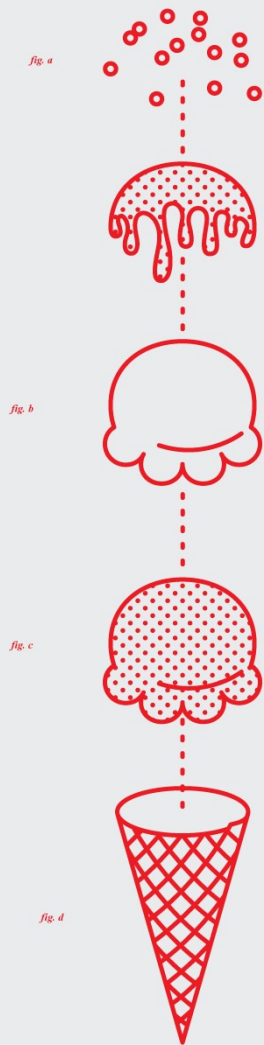
My top 3 tips

1. Forget about Excel.

2. Learn by doing. Break down your task into small chunks and try to solve them one by one.

3. Use the community knowledge.