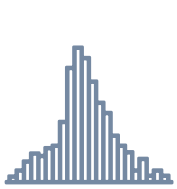
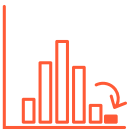
EXPLORATORY DATA ANALYSIS
= check list =

Understanding
Browse the data, columns and data types
check your domain knowledge

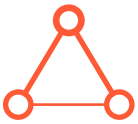
Hypothesis
what are your assumptions
ask yourself questions



Explore
look for groups, skewness, unexpected centrality and spread, re-express your data if needed: log, root,..



Clean
deal with missing values,
why are they missing?
extreme values..
are they really outliers?



Relationships
check for correlations between values
are all correlations making sense?

Back to the hypothesis
were your assumptions correct?
did you tackle the right questions?

Fine tune
keep only relevant and non-redundant plots, check that all plots are clear and self explanatory

Explain
add explanations and overviews
document your thought process..
WHY did you do all the analysis?

"The greatest value of a picture is when it forces us to notice what we never expected to see" ~John Tukey

Tereza Iofciu / @terezaif

neue fische
School and Pool for Digital Talent