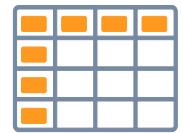




# EXPLORATORY DATA ANALYSIS

## = check list =

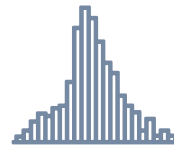


### Hypothesis

what are your assumptions  
ask yourself questions

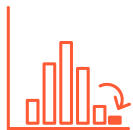


**Understanding**  
Browse the data, columns and data types  
check your domain knowledge



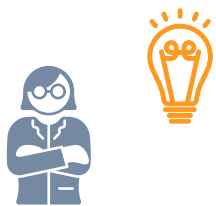
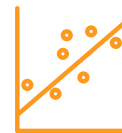
### Explore

look for groups, skewness, the 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?



### Back to the hypothesis

were your assumptions correct?  
did you tackle the right questions?

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



### Explain

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

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

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

