Data Communication Concepts
by datacamp

ADEPT technique by Kalid Azad
Analogy
Diagram
Example
Plain English
Technical definition

Be careful of acronyms and jargon

Selecting the right data
variations of data
- absolute ie difference between 2017 and 2018 sales ie total units sold
    - focusing on one variable this is a good option
- relative ie percentage variation from 2017 and 2018 ie percentage of change for each product
    - good option for multiple variables

Absolute and relative change depend on the quantity
For relative change small numbers can appear more significant than reality

Ratio
-quotient of two variables
-ie revenue per customer (total product revenue / number of customers)
Ratios help normalize values, which in turn helps compare the distribution of data of originally different scales

Aggregates
a number that gives an idea of an overall or representative value
it can be a simple total or count
ie total sales or length of campaign
*common aggregate is the mean or median

p-value
proof of statistical significance ie not due to randomness
by convention a value less than 0.05 is statistically significant
the lower it gets below 0.05 the stronger the indicator
only rejects or accepts a hypothesis
*it is not a proof of evidence

Data storytelling
  1. data
  2. narrative
  3. visuals

Data storytelling road
story> tech or not tech> data> viz> present

Pareto principle
states that the majority of outputs come from a minority of inputs
similar to the 80-20 rule

Best practices
aggregate less relevant data
approachable and engaging visuals (consider audience background)
speed and quantity of insights and how the audience receives them
less is often more

McCandless method
  1. introduce visualization by name
      1. graph headline (clear and obvious)
  2. anticipate audience's questions
  3. state insights
      1. describe graph
  4. help the audience relate

Good communication format
-key information
-engaging
-easy to understand

Stakeholders (this helps us focus on the content of our written report and/or oral presentation)
who are they> why they need to know> how are they going to use the findings> what information do they need

Requirements
how much time to present
do they need documentation to report to someone else

Presentation strategy/format
audience> content> requirements> channel

Types of reports
written reports
-explain data analysis project
-communicate findings
-standards
-give recommendations to drive change

informational
-factual information
-short
-not strict structure
-inform about facts

Analytical
-analysis (relationships/recommendations)
-varies (could be short or long)
-strict structure
-dota-driven decisions

Final report
-elements > data analysis, findings and results, and visuals
-format > long
-audience > requires detailed overview

Summary report
-key findings
-visuals
-format > short (<5 pages), summary of final report, attach link to final report
-audience does not require technical details

Report structure for analytical, final, and summary reports
  − introduction
      − purpose
      − contextual information
      − question of analysis
  − body
      − data via description and tables
      − methods
  − analysis
      − visuals
  − results
      − description and visuals

- conclusions
    - restate question
    - summarize important results
    - add recommendations

Report structure in a business context
*good method is the 1-3-25 approach:
1pg abstract > 3pg max exec summ > 25pg max details

Reproducibility
reproducing project with same parts and system

Replicability
replicating a project in a different environment (say different code or different sample)

Both these concepts are important and hold strong the virtues of Francis Bacon's s scientific method
prevents duplication of effort
build upon pre-existing work
focus on new challenges
allows for peer review
tool agnostic

Best practices for replication and reproducibility
- Keep track of how results were produced with well documented scripts and by placing comments in the code
- list all packages, libraries, and environment used
- use version control like git to easily keep track of all edits and revision and add-on's
- **never manually manipulate the data
    - save all data versioning when possible
    - at the minimum store raw data and intermediate steps
    - the story of our data transformation can help create the narrative around it
    - doing all this ensures that at all points during the process we know what is going on with the data
    - particularly need to be in control prior to data imputation
- use random seeds when working with randomness
- understand the models prediction
- cite references

Reports should be precise and concise
**avoid empty phrases

active voice for business context
passive voice for academic context

avoid redundant adverbs or adjectives
avoid run-on sentences

Example - report on credit risk
Story:
Background
      - Increase in defaulting percentage over last five years
      - Interest in predicting which customers had a high probability of default
Insight: People with more unemployment periods tend to default more
Insight: People with lower income tend to default more
Climax: Possible to predict which people is more likely to default with an accuracy of 95%
Next steps: Run a trial on a control population

Tech or not tech
Audience persona
Role: Financing Department Director
Interest: Decision on implementing an automated loan rejection system

Data:
      - Relationship between age or income and loan default
      - Percentage of customers defaulting over the next months
Statistics
-median age and income for default versus non-default customers
-to show how the number of customers defaulting changes over time, we show the percentage of change

Visuals:
Boxplot with age vs default condition
Lineplot with percent change in defaulting customers over the next months

Present:
encapsulates
Who?: Financial Department Director
Why?: important decisions ahead
Content: key findings and recommendations
Channel: email the results before the meeting

Decision made for a written report

-summary over final report d/t non-technical nature of persona
-analytical report over informational d/t need to express the analysis as well as the facts

Our summary report structure:
Introduction
  – purpose
  – contextual information
  – question of analysis
Body
  – data
  – results showing key findings
Conclusions
  – restate question
  – central insight
  – add recommendations

Planning an oral presentation
-consider its purpose
-who is the audience
-what is the message that we want to get across

What type of presentation?
informative
instructional
persuasive

What is the central message?
**After one week the audience will have forgotten over 90% of the presentation
What do we want to stick?
then work backwards from there
this should be stated in the opening statement to capture the audience's attention
advisable to be stated in one sentence
closing statement should sum up presentation and reiterate/strengthen central message

Oral presentation structure
Intro
  – provide background information
  – catch audience's attention
  – glimpse of presentation content
Methods, analysis and model outputs

Conclusions and takeaways
  – refer back to introduction
  – contains call-to-action statement and/or next steps

Building slides
no more than 3 colors
contrast text against background
use colors that will not affect color blind individuals
'sans-serif' is the pro font choice
text should support, *should not suffice without a presenter
have headlines with main point of the slide
layer slides with multiple points
if requiring a lot of text to portray point, then consider a graph instead
only one graph per slide

Presenting
write script
don't memorize
practice, practice, practice
anticipate follow-up questions
practice as if truly happening (standing, out loud, and with slides)
detect distracting patterns and stops (like like and um)
audience attention is about 5-20 minutes depending on setting
do not go expected time
talk to audience (eye contact)
empathize with audience
be interactive when possible
pace for yourself and the audience
always have a question section

Avoiding common mistakes
Always be clear with the purpose in the opening statement
sprinkle findings in as you go to contribute to the audience's participation with the story
do not give the presentation as a monologue