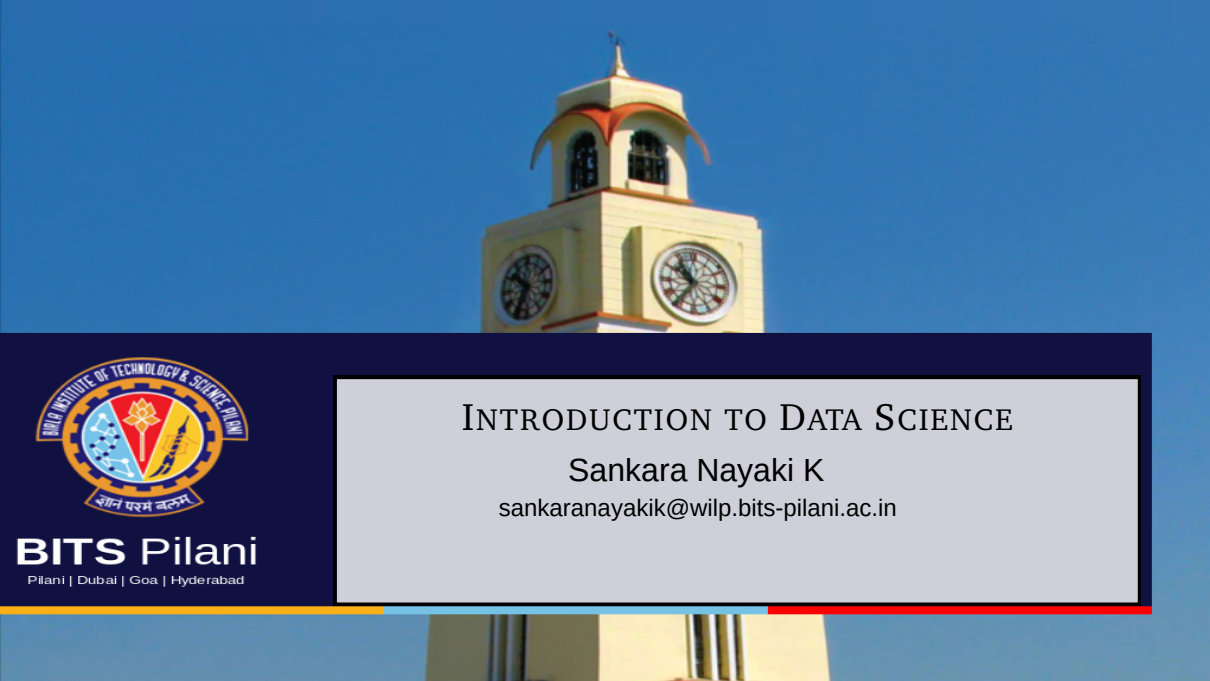


BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Data Science

Sankara Nayaki K
sankaranayakik@wilp.bits-pilani.ac.in

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

References:

- Introducing Data Science by Cielen, Meysman and Ali
- Storytelling with Data by Cole Nussbaumer Knaflic; Wiley
- Introduction to Data Mining by Tan, Steinbach and Vipin Kumar
- The Art of Data Science by Roger D Peng and Elizabeth Matsui
- Python Data Science Handbook: Essential tools for working with data by Jake VanderPlas

Table of Contents

Course Handout

M1 Introduction to Data Science
M2 Data Analytics
M3 Data Science Process
M4 Data Science Teams
M5 Data & Data Models
M6 Data Wrangling & Feature Engineering
M7 Data Visualization
M8 Storytelling with Data
M9 Ethics for Data Science

Table of Contents

Magic of Story

- A good story grabs your attention and takes you on a journey, evoking an emotional response. In the middle of it, you find yourself not wanting to turn away or put it down. After finishing it—a day, a week, or even a month later—you could easily describe it to a friend.

Story

- Story is a time-tested structure.
- Story has a clear beginning, middle, and end.
- This can be setup, conflict, and resolution.
- Stories start with balance. Then some event happens, that upsets the balance. The resulting struggle, conflict, and suspense are critical components of the story.

The Beginning

- Introduce the plot, build the context for your audience.
- The setting, main character, unresolved state of affairs, and desired outcome.

1. The setting: When and where does the story take place?
2. The main character: Who is driving the action? (framed in terms of your audience)
3. The imbalance: Why is it necessary, what has changed?
4. The balance: What do you want to see happen?
5. The solution: How will you bring about the changes?

The Middle

- “what could be”
- The goal is to convince the audience of the need for action.
- Identify what motivates the audience.
- Consider framing your story and the need for action in terms of this.
- Use data to strengthen the story and integrate it to make sense.

The Middle

1. Further develop the situation or problem by covering relevant background.
2. Incorporate external context or comparison points.
3. Give examples that illustrate the issue.
4. Include data that demonstrates the problem.
5. Articulate what will happen if no action is taken or no change is made.
6. Discuss potential options for addressing the problem.
7. Illustrate the benefits of your recommended solution.
8. Make it clear to your audience why they are in a unique position to make a decision or drive action.

The End

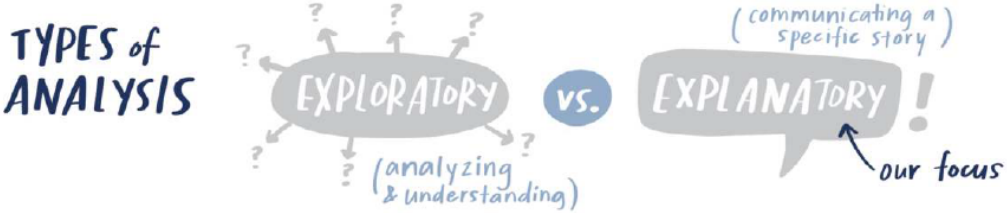
- Call to action.
- The audience is totally clear on what you want them to do with the new understanding or knowledge that you've imparted to them.

The Narrative Flow

- Your story must have an order to it.

1. Are they a busy audience who will appreciate if you lead with what you want from them?
2. Are they a new audience, with whom you need to establish credibility?
3. Do they care about your process or just want the answer?
4. Is it a collaborative process through which you need their input?
5. Are you asking them to make a decision or take an action?
6. How can you best convince them to act in the way you want them to?

Importance of Context



Storytelling with Data by Cole Nussbaumer Knaflic

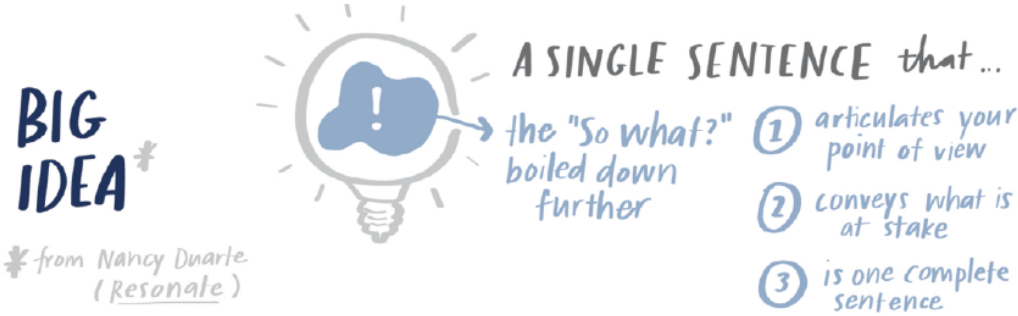
Importance of Context

Storytelling with Data by Cole Nussbaumer Knaflic

Storytelling

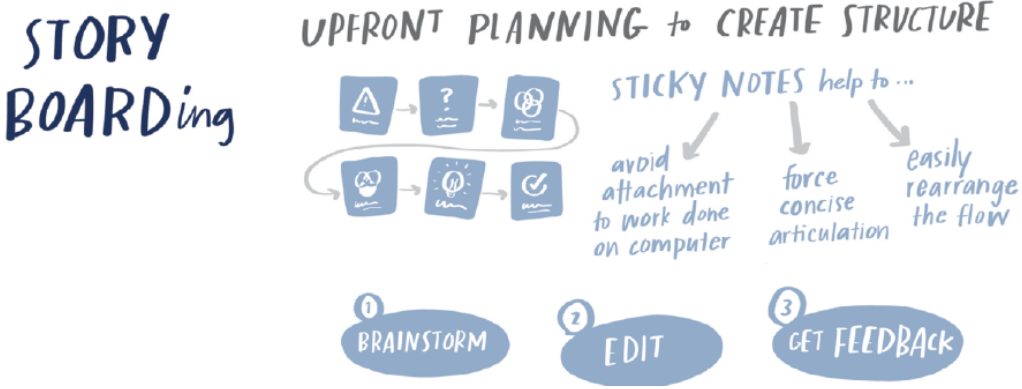
Storytelling with Data by Cole Nussbaumer Knaflic

Storytelling



Storytelling with Data by Cole Nussbaumer Knaflic

Storytelling



Storytelling with Data by Cole Nussbaumer Knaflic

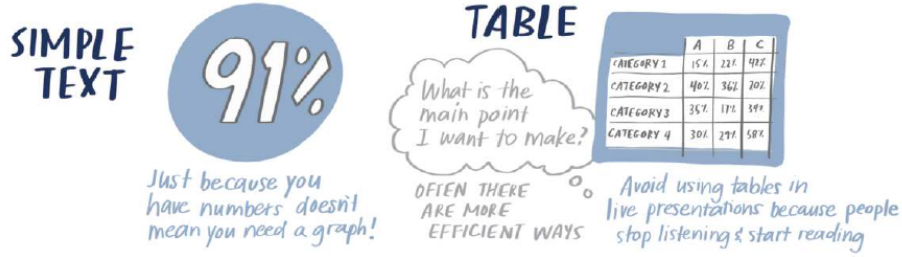
Storyboarding

Storytelling with Data by Cole Nussbaumer Knaflic

Storyboarding

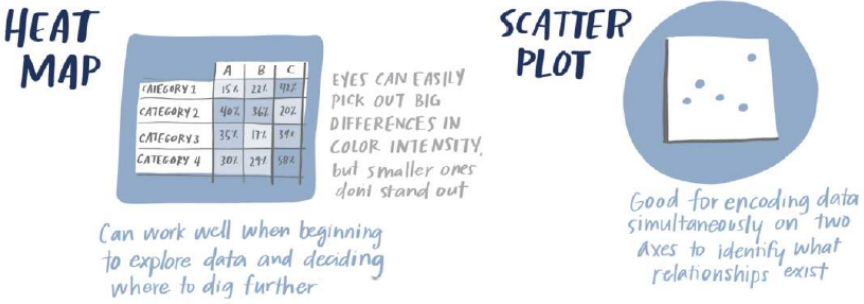
Storytelling with Data by Cole Nussbaumer Knaflic

Choosing an Effective Visual



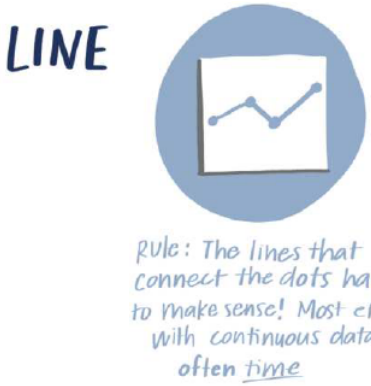
Storytelling with Data by Cole Nussbaumer Knaflic

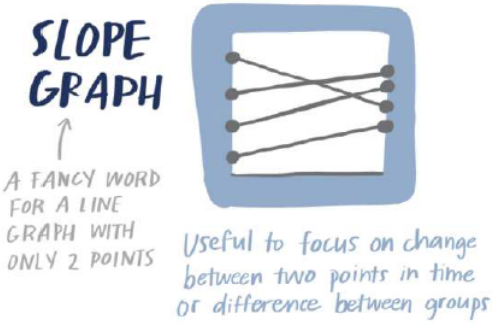
Choosing an Effective Visual



Storytelling with Data by Cole Nussbaumer Knaflic

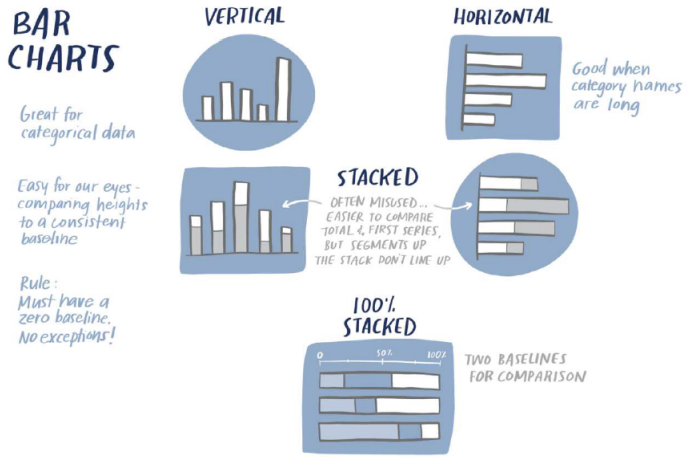
Choosing an Effective Visual





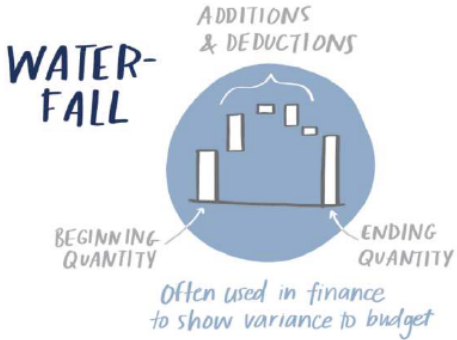
Storytelling with Data by Cole Nussbaumer Knaflic

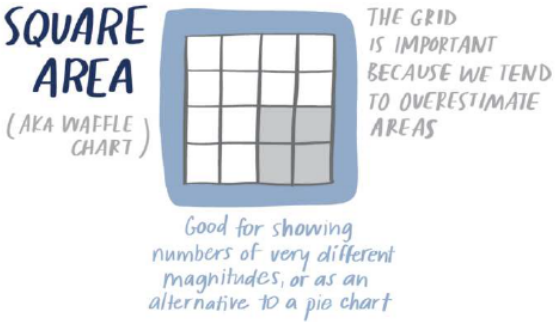
Choosing an Effective Visual



Storytelling with Data by Cole Nussbaumer Knaflic

Choosing an Effective Visual





Storytelling with Data by Cole Nussbaumer Knaflic

Improving the Table

New client tier share

Tier	# of Accounts	% Accounts	Revenue ($M)	% Revenue
A	77	7.08%	$4.68	25%
A+	19	1.75%	$3.93	21%
B	338	31.07%	$5.98	32%
C	425	39.06%	$2.81	15%
D	24	2.21%	$0.37	2%

Storytelling with Data by Cole Nussbaumer Knaflic

Improving the Table: Option

New client tier share

TIER	ACCOUNTS		REVENUE	
	#	% OF TOT	$M	% OF TOT
A+	19	2%	$3.9	21%
A	77	7%	$4.7	25%
B	338	31%	$6.0	32%
C	425	39%	$2.8	15%
D	24	2%	$0.4	2%
All other	205	19%	$0.9	5%
TOTAL	1,088	100%	$18.7	100%

Storytelling with Data by Cole Nussbaumer Knaflic

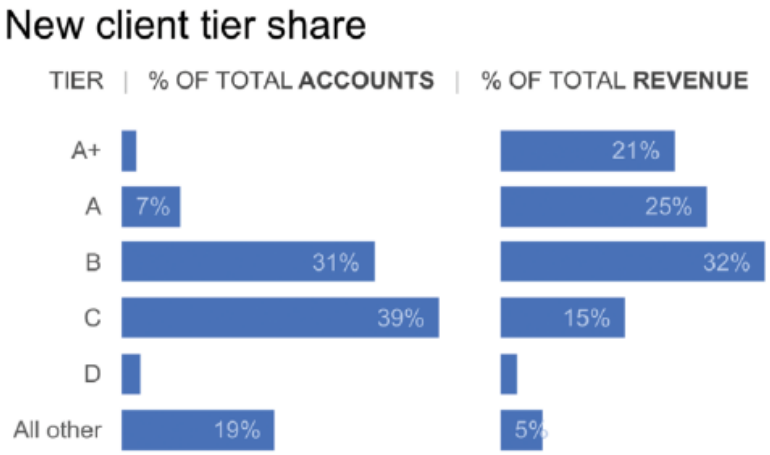
Improving the Table: Option

New client tier share

TIER	ACCOUNTS #	ACCOUNTS % OF TOT	REVENUE $M	REVENUE % OF TOT
A+	19		$3.9	
A	77		$4.7	
B	338		$6.0	
C	425		$2.8	
D	24		$0.4	
All other	205		$0.9	
TOTAL	1,088	100%	$18.7	100%

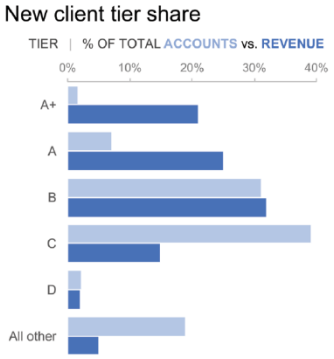
Storytelling with Data by Cole Nussbaumer Knaflic

Improving the Table: Option



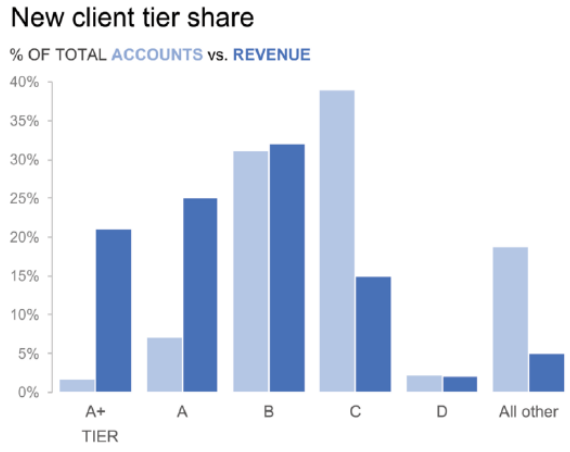
Storytelling with Data by Cole Nussbaumer Knaflic

Improving the Table: Option



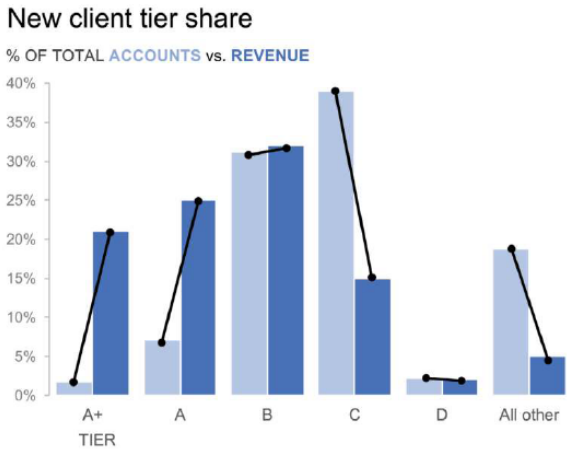
Storytelling with Data by Cole Nussbaumer Knaflic

Improving the Table: Option



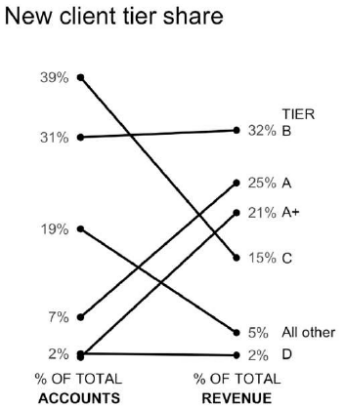
Storytelling with Data by Cole Nussbaumer Knaflic

Improving the Table: Option



Storytelling with Data by Cole Nussbaumer Knaflic

Improving the Table: Option



Storytelling with Data by Cole Nussbaumer Knaflic

Visualize

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

Storytelling with Data by Cole Nussbaumer Knaflic

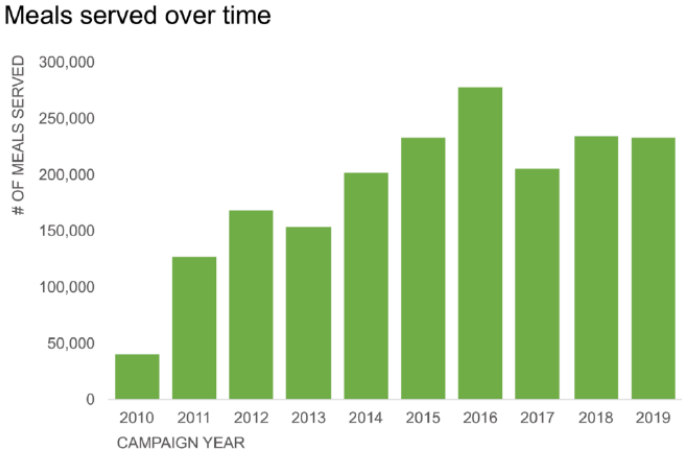
Visualize: Options

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

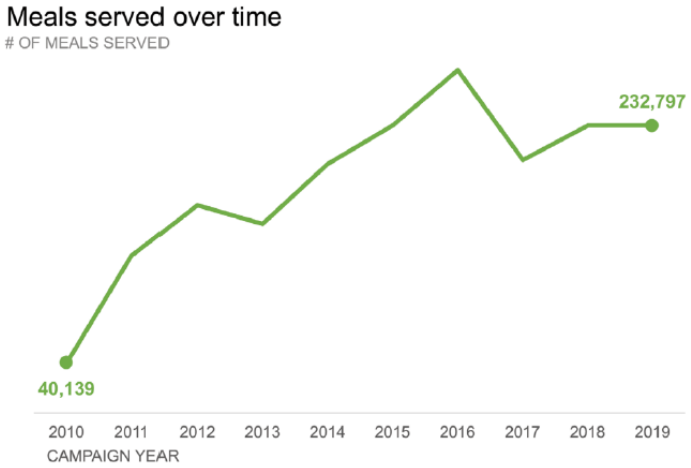
Storytelling with Data by Cole Nussbaumer Knaflic

Visualize: Options



Storytelling with Data by Cole Nussbaumer Knaflic

Visualize: Options



Storytelling with Data by Cole Nussbaumer Knaflic

Correct the Visualization

Storytelling with Data by Cole Nussbaumer Knaflic

THANK YOU