Introduction to Data Science

Data Visualization

SANKARA NAYAKI K

sankaranayaki@wilp.bits-pilani.ac.in

Disclaimer and Acknowledgement

- The content for these slides has been obtained from books and various other source on the Internet
- I here by acknowledge all the contributors for their material and inputs.
- I have provided source information wherever necessary
- I have added and modified the content to suit the requirements of the course

Data Visualizations

- Data Need for visualization
- Exploratory vs Explanatory Analysis

- Tables , Axis based Visualization and Statistical Plots
- Lessons in Data Visualization Design
- The Data Visualization Design Process
- Stories and Dashboards

Case Studies

CASE STUDY - #1

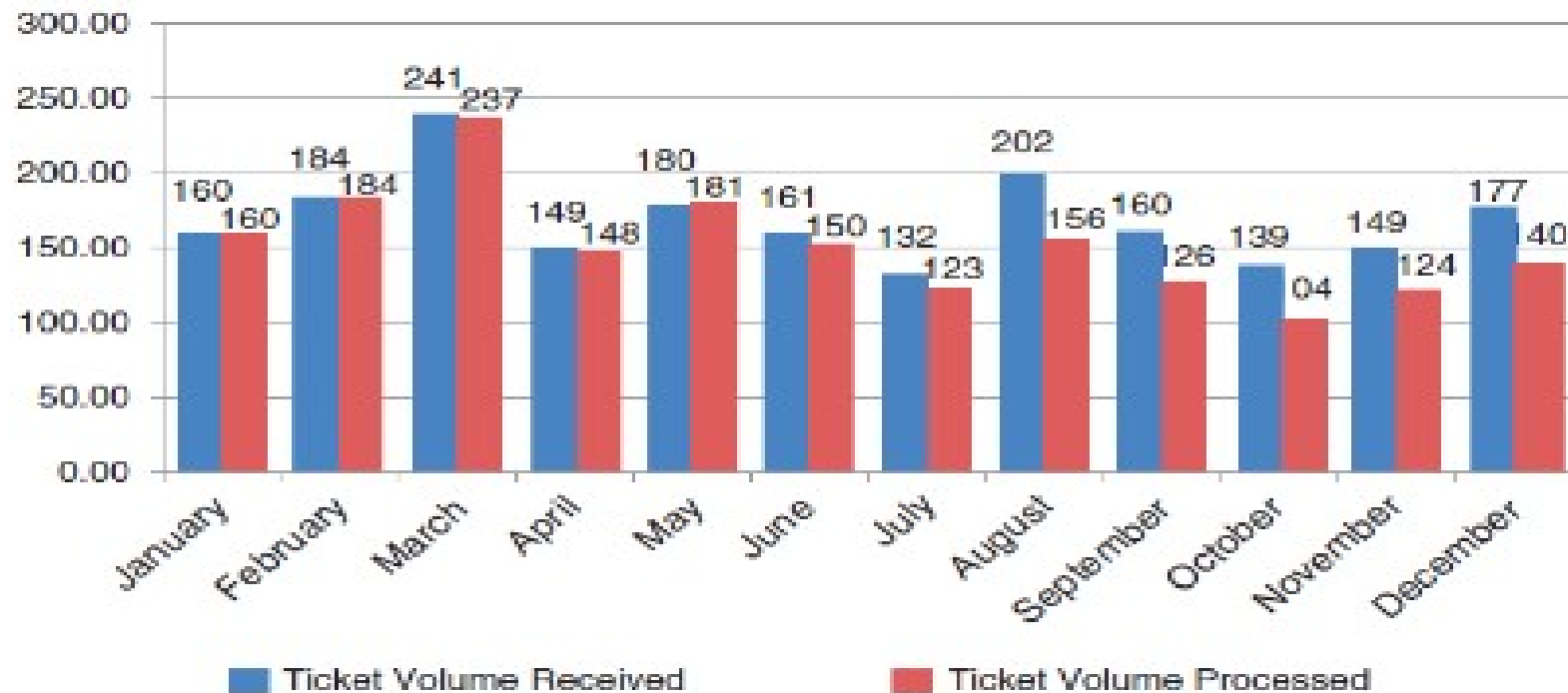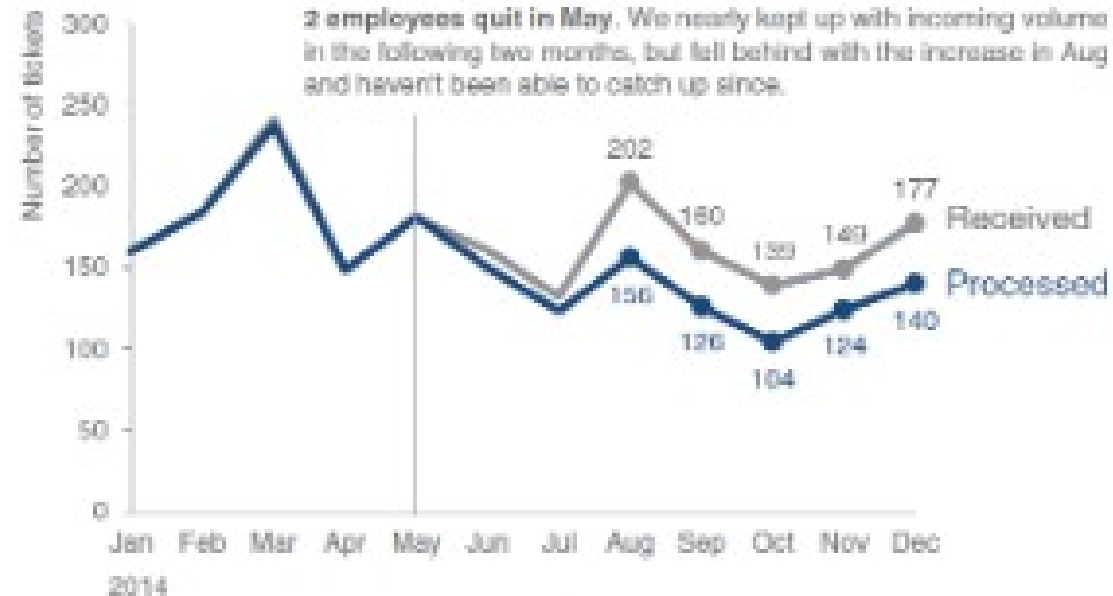
Source: Storytelling with Data by knaflic

innovate   achieve   lead

Please approve the hire of 2 FTEs
to backfill those who quit in the past year

Ticket volume over time

2 employees quit in May. We nearly kept up with incoming volume in the following two months, but fell behind with the increase in Aug and haven't been able to catch up since.

Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.
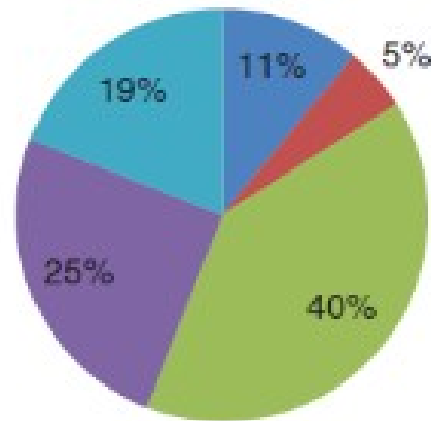
Source: Storytelling with Data by knaflic

CASE STUDY - #2

Survey Results

PRE: How do you feel about doing science?

Bored ■ Not great ■ OK ■ Kind of interested ■ Excited

- 11%
- 5%
- 40%
- 25%
- 19%

POST: How do you feel about doing science?

Bored ■ Not great ■ OK ■ Kind of interested ■ Excited

- 12%
- 6%
- 14%
- 30%
- 38%

Source: Storytelling with Data by knaflic
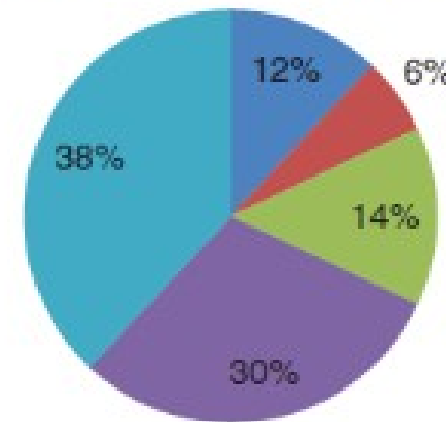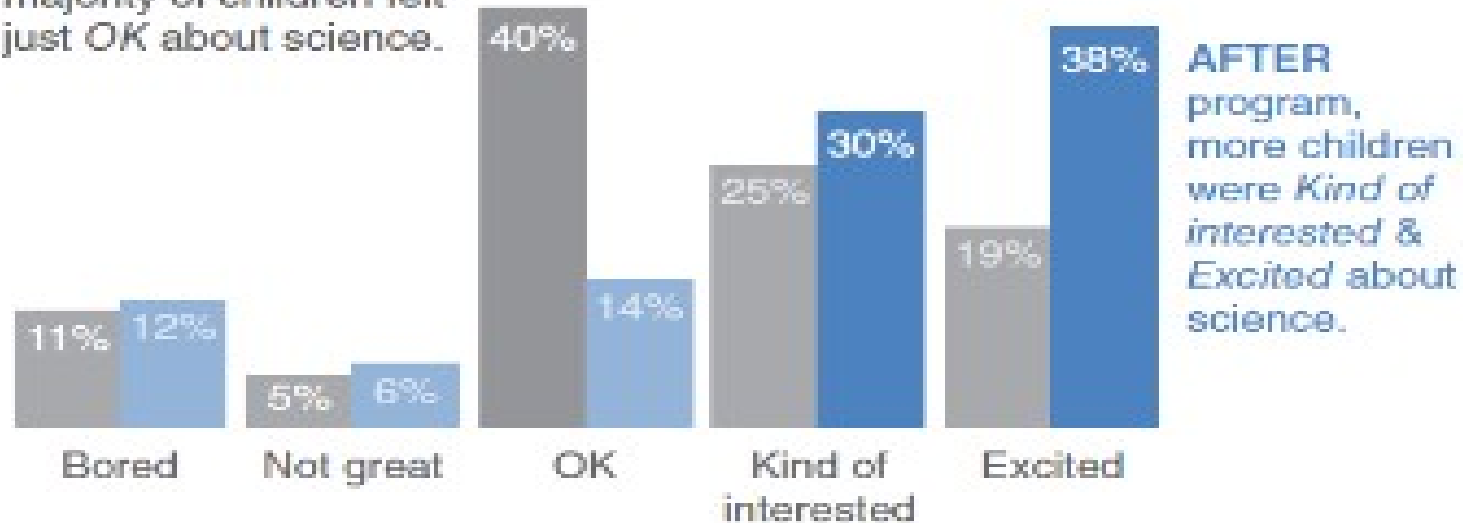
Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just *OK* about science.

AFTER program, more children were *Kind of interested* & *Excited* about science.

11% 12% | 5% 6% | 40% 14% | 25% 30% | 19% 38%

Bored | Not great | OK | Kind of interested | Excited

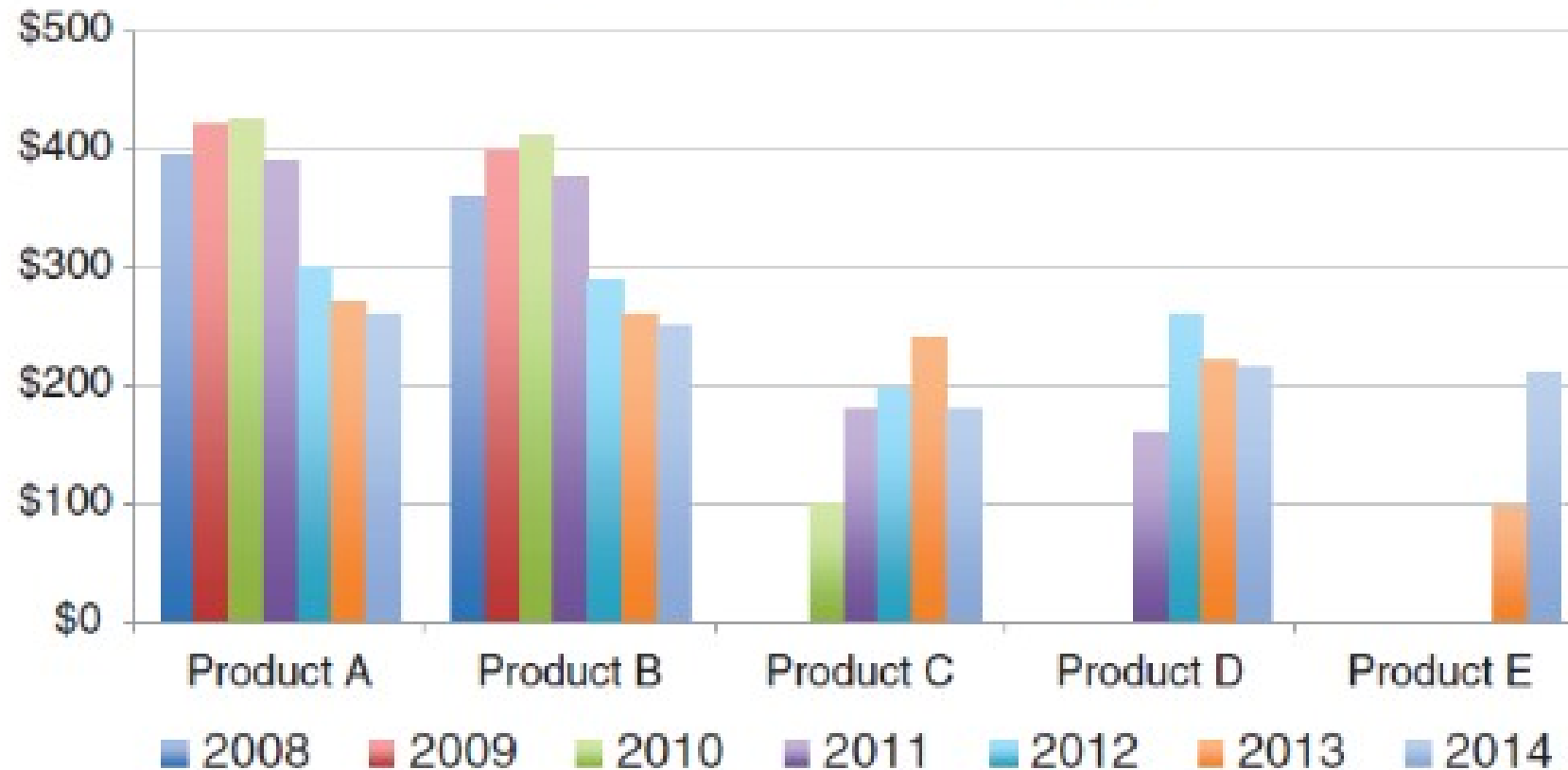Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

Source: Storytelling with Data by knaflic

CASE STUDY - #3

Average Retail Product Price per Year
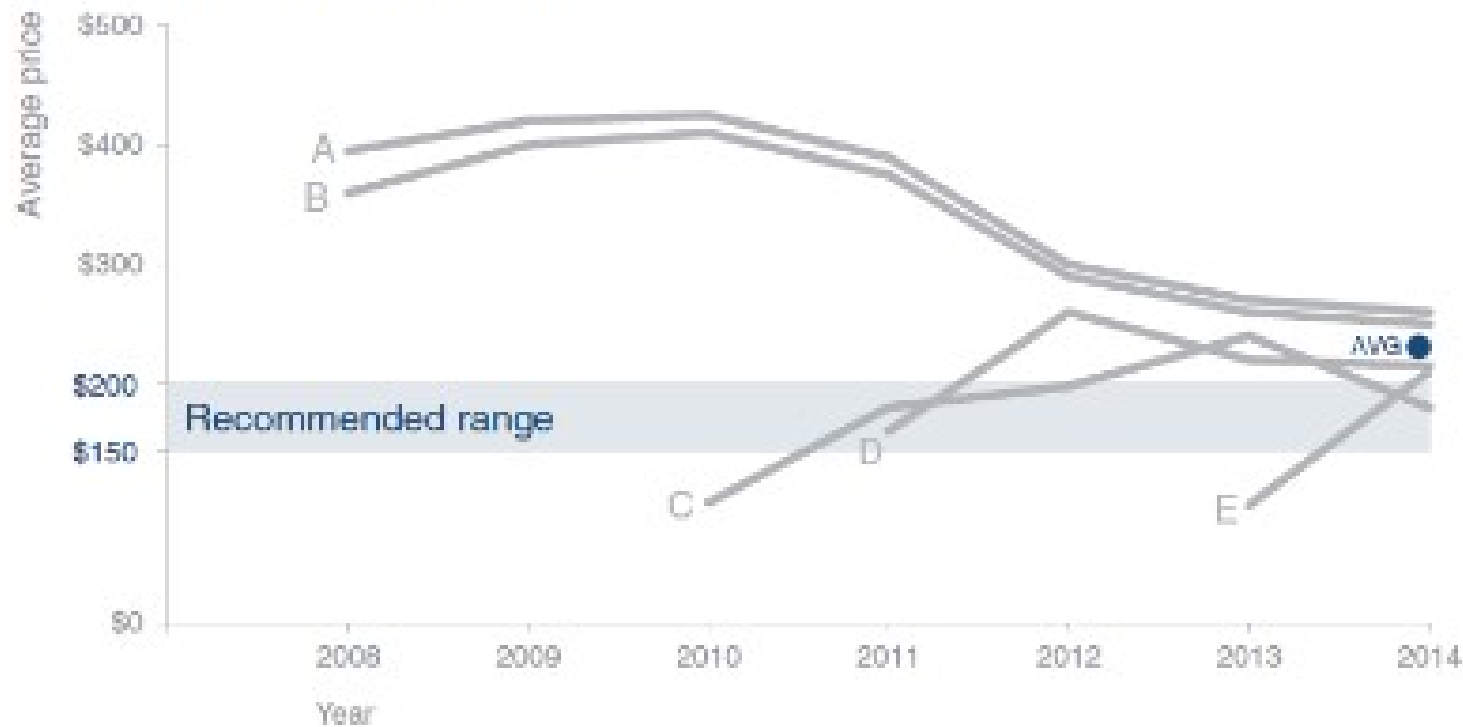
innovate   achieve   lead

Need for Visualization

innovate   achieve   lead

Need for Visualization

Introduction

- What?
  - Graphical / Visual representation of data

- Why?
  - Way to identify patterns, trends and outliers in data
  - Helps is making data-driven decisions

- "The purpose of visualization is insight, not just the picture."
  - Data visualization pioneer, Ben Shneiderman

Why do we visualize quantitative data?

- We visualize quantitative data to perform three fundamental tasks in an effort to achieve three essential goals:

Goal	Task
Discovery	Exploration: Searching for significant facts
Understanding	Sensemaking: Examining and making sense of data
Informed decisions	Communication: Conveying information to others

- Based on these goals and tasks, data visualization is defined as:
  - "Data visualization is the use of visual representations to explore, make sense of, and communicate data"

Source: Visual Business Intelligence by Stephen Few

Myths about data visualization

- Why should we use graphical displays (rather than other forms such as tables and audio) to communicate and convey the information to others?

- There is only one good reason to express quantitative data visually:
  - Some features can be best perceived and understood when quantitative data displayed graphically

- This is because of the ways our brains work

- Vision is by far our dominant sense

- We have evolved to perform many data sensing and processing tasks visually

- Data exploration, sensemaking, and communication should always involve an intimate collaboration between seeing and thinking (i.e., visual thinking).

Need for Visualization

Myths about data visualization

- Myth #1: We visualize data because some people are visual learners.
  - True, but everyone benefits from data visualization, whether they consider themselves visual learners or not, including those who prefer numbers.
- Myth #2: We visualize data for people who have difficulty understanding numbers.
  - True, but even the brightest mathematicians benefit from seeing quantitative information displayed visually. Data visualization is not a dumbed-down expression of quantitative concepts.
- Myth #3: We visualize data to grab people's attention with eye-catching but inevitably less informative displays.
  - It isn't necessary to sacrifice content in lieu of appearance. Data can always be displayed in ways that are optimally informative, pleasing to the eye, and engaging. To engage with a data display without being well informed of something useful is a waste.
- Myth #4: The best data visualizers are those who have been trained in graphic arts.
  - While training in graphic arts can be useful, it is much more important to understand the data and be trained in visual thinking and communication.
- Myth #5: Graphics provide the best means of telling stories contained in data.
  - While it is true that graphics are often useful and sometimes even essential for data-based storytelling, it isn't storytelling itself that demands graphics. Much of storytelling is best expressed in words and numbers rather than images. Graphics are useful for storytelling because some features of data are best understood by our brains when they're presented visually.

Source: Visual Business Intelligence by Stephen Few

innovate   achieve   lead

Why do we really visualize data?

- Because humans can perceive particular quantitative features and perform particular quantitative tasks most effectively when the data is expressed graphically

- Visual data processing provides optimal support for the following:
  - Seeing the big picture
  - Easily and rapidly comparing values
  - Seeing patterns among values
  - Comparing patterns

Source: Visual Business Intelligence by Stephen Few

Example

• Seeing patterns among values

Sales

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493
International	574	636	673	593	644	679	593	139	599	583	602	690
	$2,557	$2,979	$3,266	$2,876	$3,218	$3,517	$2,975	$2,773	$3,537	$3,322	$3,585	$4,183
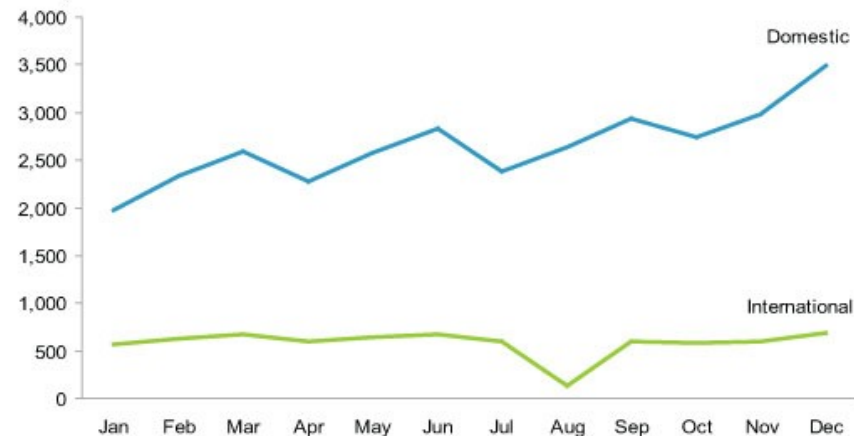
Source: Visual Business Intelligence by Stephen Few

Types of Visualizations

- Static Visualizations

- Interactive/Dynamic Visualizations

Types of Visualizations

- Static Visualizations
  - They are commonly seen as infographics posted on the web or printed as handouts
  - Used in presentations, documents, etc.
  - Requires careful design as its meant for offline viewing
  - User can not adjust the views
  - Usually focused on a specific data story, users can't go beyond a single view to explore additional stories beyond what's in front of them
  - The story is specifically captured in an engaging single page layout.

Types of Visualizations

- Interactive/Dynamic Visualizations
  - They are commonly seen on the web only as applications
  - Users can select specific data points to build a visualized story of their choosing
  - Used for exploratory data analysis
  - Meant for live/online interactions
  - Has more viewing options
  - User gains more control over the display
  - These visualizations allow the user to be part of the data visualization process by building a story of their choosing

Types of Visualizations

- Common/ General Types of Data Visualization
  - Charts
  - Tables
  - Graphs
  - Maps
  - Infographics
  - Dashboards

- Specific Examples of Methods to Visualize Data

- Area Chart
- Bar Chart
- Box-and-whisker Plots
- Bubble Cloud
- Bullet Graph
- Cartogram
- Circle View
- Dot Distribution Map

- Gantt Chart
- Heat Map
- Highlight Table
- Histogram
- Matrix
- Network
- Polar Area
- Radial Tree
- Scatter Plot (2D or 3D)

- Streamgraph
- Text Tables
- Timeline
- Treemap
- Wedge Stack Graph
- Word Cloud
- Any combination of the above

Exploratory Vs. Explanatory Analysis

- Exploratory Analysis
  - Required to become familiar with data
  - Involves digging through the data
  - Helps in finding trends and relationships w.r.t. specific goals
  - It's like hunting for pearls in oysters
  - The purpose is to understand the data and figure out what might be noteworthy or interesting to highlight to others
    - E.g., hunting for pearls in oysters
  - We might have to open 100 oysters (test 100 different hypotheses or look at the data in 100 different ways) to find perhaps two pearls.

Source: Storytelling with Data by knaflic

Exploratory Vs. Explanatory Analysis

- Explanatory Analysis
  - It's about communicating our findings
  - It's about telling a story with data
  - We need to be in the explanatory space when we want to communicate the results of our analysis to our audience
  - At this stage, we should have a specific story to tell or a specific thing we want to explain
    - E.g., probably about those two pearls.

Source: Storytelling with Data by knaflic

Data Visualization – Context

Agenda

- Under this topic, we will focus on the following aspects:
  - Understanding the context
  - Effective storytelling strategies

Success criteria

- "Success in data visualization does not start with data visualization itself." -- Cole Knaflic

- …understanding the context sets solid foundation for data visualization creation

Source: Storytelling with Data by knaflic

Who, What, & How

- Consider three things in explanatory analysis and data visualization:
  - Who
    - First, to whom are we communicating?
    - Who is our audience and how they perceive us?
    - It is important to identify common ground so that we can ensure they hear our message
  - What
    - Second, what do we want our audience to know or do?
    - We should be clear about how our audience should act and take into account our communication
  - How
    - Next, how can we use data to help make your point?

Source: Storytelling with Data by knaflic

Who – Our Audience

- Our audience
  - Being clear about our audience helps us communicate better
Knowing them places us in a better position for communication
    - We need to be specific while identifying our audience
    - Avoid general audiences, such as:
      - "internal and external stakeholders" or
      - "anyone who might be interested"
    - If we have audience with disparate needs, we cannot communicate effectively to anyone
      - We may need to create different communications for different audiences

Source: Storytelling with Data by knaflic

Who - We

- We, the presenters
  - Key considerations to be given when determining how to structure our communication:
  - What is our relationship with our audience?
  - How do they perceive us?
  - Are we encountering with our audience for the first time through this communication or do we already have an established relationship?
  - Do they trust us as an expert, or do we need to work to establish credibility?
  - Whether and when to use data, and may impact the order and flow of the overall story we aim to tell

Source: Storytelling with Data by knaflic

What

- What do we want our audience to know or to do?
  - Action
  - Mechanism
  - Tone

What – Action

- Make what we communicate relevant for our audience

- It helps us in ensuring our audience care about what we have to say

- Our objective should be such that we always want our audience to know or do something

- Sometimes, we may not concisely articulate what we need to communicate

- This could be because we believe our audience knows better than us
  - This assumption is false

- We are the subject matter experts here because we are the one analyzing the data

- This puts us in a unique position to interpret the data and help lead people to understanding and take action

- If no action recommendation possible / feasible, then encourage discussion towards one

- In general, the presenter should take a more confident stance when it comes to making specific observations and recommendations based on their analysis

What – Mechanism

- How will you communicate to your audience?

- Communication mechanism can be placed on a continuum of:
  - Live presentation on one end to a written document/email on the other

- Broadly, three levels of communication mechanisms
  - Live presentation
  - Written document or email
  - Slideument

- The method we choose to communicate to our audience determines:
  - the level of control we will have over how the audience takes in the information and
  - the level of detail that needs to be explicit

Source: Storytelling with Data by knaflic
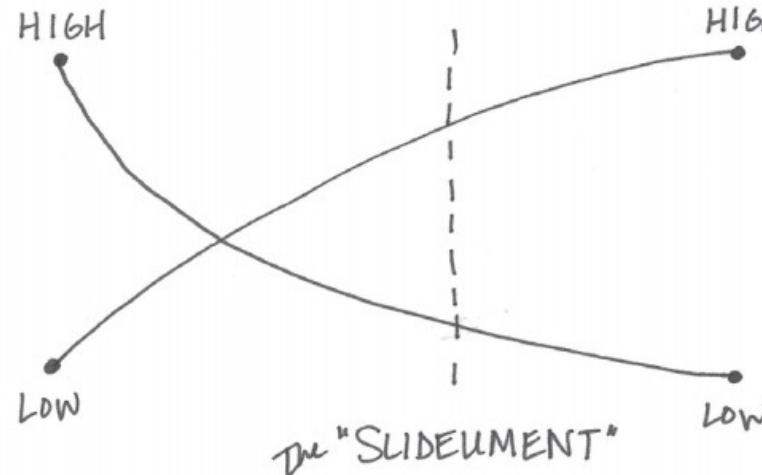
What – Mechanism

Live
Presentation

Written
Doc/Email

amount of
CONTROL
you have

level of
DETAIL
needed

- Consider:
  - the level of control we have over how the information is consumed
  - the amount of detail needed

  at either end of the spectrum.

What – Mechanism

- Live Presentation
  - We (the presenter) are in full control
  - We determine what the audience sees and when they see it
  - We can respond to visual cues to:
    - speed up, slow down, or go into a particular point in more or less detail
  - Not all of the detail needs to be directly in the slide deck
  - We, the subject matter expert, are there to answer any questions during the presentation

Live Presentation

Written Doc/Email

amount of CONTROL you have

level of DETAIL needed

HIGH

HIGH

LOW

LOW

The "SLIDEUMENT"

Source: Storytelling with Data by knaflic
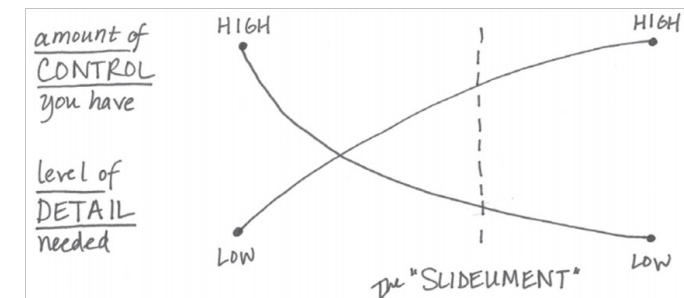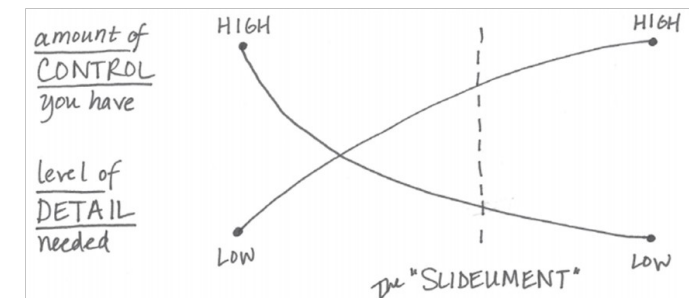
What – Mechanism

- Written Document/Email
  - We (the creator of the document or email) have less control
  - The audience is in control of how they consume the information
  - The level of detail that is needed here is typically higher because we aren't there to see and respond to your audience's cues
  - Rather, the document will need to directly address more of the potential questions

Live Presentation

Written Doc/Email

What – Mechanism

- Slideument
  - In an ideal world (due to time and other constraints), we create a single document (slideument), solve both of these needs
  - This poses some challenges because of the diverse needs it is meant to satisfy
  - It is important to identify the primary communication means we'll be leveraging:
    - live presentation, written document, or something else.

Live Presentation

Written Doc/Email

amount of CONTROL you have

level of DETAIL needed

The "SLIDEUMENT"
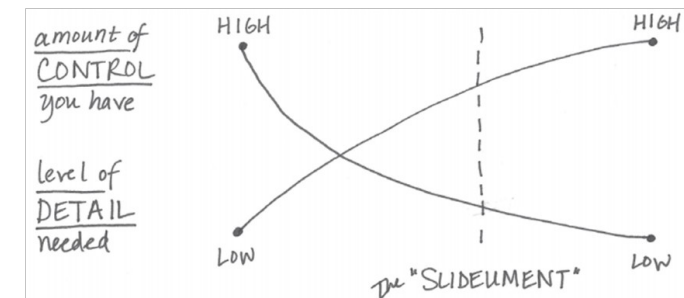
What – Tone

- What tone do we want our communication to set?
  - Are we celebrating a success?
  - Are we trying to light a fire to drive action?
  - Is the topic light-hearted or serious?

- Think about the general tone that we want to establish when we set out on the data visualization path.

Source: Storytelling with Data by knaflic

Data Visualization – Context

How – Data

- Finally, after having a clear idea about our audience and what we need them to know or do
  - We can ask the question: What data is available that will help make our point?

- What data is available?

- How can we use data to help make our point?
  - Data becomes supporting evidence for our story

Context by Example

Scenario

- You are advertising media analyst made responsible for providing recommendation on media spend of product

- Who:
  - The marketing team that allocates funding for media advertisement for a product

- What:
  - The current advertising campaign went well on TV but find very limited success in print media

- How:
  - Illustrate success with data available through analysis of spends and product revenues

Source: Storytelling with Data by knaflic

Question to ponder

- What background information is relevant or essential?

- Audience
  - Who is the audience or decision maker?
  - What do we know about them?
  - What biases does our audience have that might make them supportive of or resistant to our message?

- Data
  - What data is available that would strengthen our case?
  - Is our audience familiar with this data, or is it new?

- Risks
  - What factors could weaken our case and do we need to proactively address them?

- What would a successful outcome look like?

- If you only had a limited amount of time to tell your audience,
  - what they need to know, what would you say?

Context by Example

Storytelling Strategies

- Concepts for successful storytelling activity
  - The 3-minute Story
  - Big Idea
  - Storyboarding

Storytelling Strategies

- "I would have written a shorter letter, but I did not have the time."
  - -- Blaise Pascal

- ...boil the overall communication down to a single paragraph or to a single, most important message

Storytelling Strategies

- 3-minute story
  - Within 3 minutes we have to tell audience, what they need to know
  - This is a great way to ensure we are clear on and can articulate the story we want to tell
  - Removes dependence from supporting material like visuals, slides, etc., in our presentations
  - Helps in a situation where we have to give elevator speech, or if our half-hour agenda is cut short to ten minutes, or to five
  - We need to know exactly what data is saying
  - Being concise is more challenging than being verbose

Source: Storytelling with Data by knaflic

Storytelling Strategies

- Big Idea
  - Boils the so-what down even further: to a single sentence
  - Per Nancy Duarte (in her book Resonate, 2010), big Idea has three components:
    - It must articulate your unique point of view;
    - It must convey what's at stake; and
    - It must be a complete sentence

Source: Storytelling with Data by knaflic

Storytelling Strategies

- Storyboarding
  - Storyboarding helps ensure the communication we craft is on point
  - A storyboard establishes a structure for our communication
  - It is a visual outline of the content we plan to create
  - It can be subject to change as we work through the details
  - When possible, get acceptance from your client or stakeholder at this step.
    - It will help ensure that what you're planning is in line with the need

Storytelling Strategies

- Storyboarding
  - Advice
    - Don't start with presentation software
      - It is too easy to go into slide-generating mode without thinking about how the pieces fit together
        - ✓ In the end, we will create a massive presentation deck that says nothing effectively
      - As as we start creating content via our computer, we will form an attachment to it, which makes us resistant to making changes or deleting slides
    - Use a whiteboard, Post-it notes, or plain paper
      - It's much easier to trash or recycle a piece of paper or a Post-it note without developing much attachment towards it

Context by Example

Storyboarding Examples

Issue

Media spends for a newly launched product has not produced desired effect on the increase in sales of the product

Demonstrate Issue

Show media spending over various channels for last 2 months

Ideas for overcoming issues

Identify new media channels
Analyze the effectiveness of established media channels

Describe Goals

Show before and after survey data to demonstrate success of the program

Recommendation

Increase media spend in TV channels
Decrease/Stop media spend in print media channels

Source: Storytelling with Data by knaflic

Recap

- Establishing context
  - Who
    - Audience and Us
  - What
    - Action, Mechanism, Tone
  - How
    - Data
- Story telling strategies
  - The 3-minute story
  - Big Idea
  - Storyboarding

Source: Storytelling with Data by knaflic

Thank You!