

Colors

Aesthetics

Flow

Tableau

Shapes

Charts

Data Visualization /

Lecture 1

Tableau Intro

File Formats

Data Extract	Workbook	Packaged Workbook
.hyper	.tub	.tubx

Process

- (1) Connect To Data Source (file/server etc.)
- (2) Review and Transform Data
 - (2.1) Data Types of columns
 - (2.2) Rename Columns
 - (2.3) Measures
- (3) Visualize
 - (3.1) Plot preliminary
 - (3.2) Use 'Marks' Pane to manipulate visual
 - (3.2.1) sequential; colored dark \rightarrow light (only + or -)
 - (3.2.2) diverging; colored dark \rightarrow light (+ AND -)
 - (3.2.3) categorical; colored by category!

Why we visualize...

There are few rules & no right answers
We create stories by thoughtfull design!

Definition

Representation & Presentation of data
to facilitate understanding. (Kirk 2016)

Avoid **data demonstration**, instead do
interesting interpretations!

Why?

- (1) a number alone says nothing
- (2) it's intuitive to understand
- (3) discover trends/patterns
- (4) enlighten people with data overview

[infowetrust.com/scroll/
viz.wtf](http://infowetrust.com/scroll/viz.wtf)

Vocabulary & Framework

5 qualities

Truthful Beautiful Functional Insightful

if you have all 4, you will achieve
enlightening which can convince people!

Visual Families :

Deviation	Correlation	Ranking	Distribution	Change over Time	Magnitude	Part-to-whole	Spatial	Flow
Example FT case Emphasize variation (X) has a trend (Y) and a deviation (Z) from the trend. This is the target and the goal of the chart.	Example FT case Show the relationship between two variables. You will then emphasize their ranking (order).	Example FT case Use where items position in an order. It's often used to show the relationship between two variables. Don't forget to emphasize the ranking.	Example FT case Show where in a dataset and how often each value was used. The distribution is often used to describe a set of data to provide valuable context.	Example FT case One emphasizes changing trends. These trends are often used to describe a set of data. Don't forget to emphasize the change over time.	Example FT case Show size or extremes. These can be used to emphasize the magnitude of the data. Don't forget to emphasize the size or extremes.	Example FT case Focus on a single entity can be broken down into its parts. These parts are often used to describe the whole. Don't forget to emphasize the part-to-whole.	Example FT case Focus on a single entity, company, organization, industry in general.	Example FT case Show the master volume or intensity of something. These can be used to describe the flow of something. Don't forget to emphasize the flow.
Dot plot A simple standard bar chart showing the range and spread of data.	Scatterplot The idealized way to show the relationship between two continuous variables. Don't forget to emphasize the correlation.	Rankline Standard bar chart where values are ranked from highest to lowest.	Histogram The idealized way to show the distribution of data. Keep the bins narrow enough so that the distribution is clear. Don't forget to emphasize the distribution.	Line The idealized way to show the change of data over time. Don't forget to emphasize the change over time.	Bar See above. Don't forget the bars are not always vertical. They can also be horizontal.	Stacked bar A simple way of showing the total value of multiple components. Don't forget to emphasize the part-to-whole.	Dot map The idealized way to show the spatial distribution of data. Don't forget to emphasize the spatial.	Timeline Shows changes in flow, movement, or intensity over time. Don't forget to emphasize the flow.
Gauge chart Best for presenting numerical results and comparing them against a target.	Box plot A good way of showing the distribution of data. Don't forget to emphasize the range and outliers.	Ordinal values See above.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Line Good for showing the change of data over time. Don't forget to emphasize the change over time.	Bar See above. Don't forget the bars are not always vertical. They can also be horizontal.	Radial sunburst A simple way of showing the hierarchical nature of data. Don't forget to emphasize the radial sunburst.	Proportional symbol The idealized way to show the spatial distribution of data. Don't forget to emphasize the spatial.	Network Used for showing the connections between entities. Don't forget to emphasize the network.
Scatterplot matrix Shows a grid where each cell contains a scatterplot comparing a third variable.	Generalized scatterplot A good way of showing the relationship between two variables. Don't forget to emphasize the correlation.	Ordinal proportional symbol Good for showing the relationship between two variables. Don't forget to emphasize the correlation.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Area chart Used with Line - Good for showing the change of data over time. Don't forget to emphasize the area.	Bar chart A good way of showing the total value of multiple components. Don't forget to emphasize the part-to-whole.	Treemap Used for hierarchical visualization. Don't forget to emphasize the hierarchy.	Dot map For drawing unstructured data. Don't forget to emphasize the dot map.	Cloud A circular field diagram showing the flow of information. Don't forget to emphasize the cloud.
Surprisational heat map The shaded area of a heatmap shows a distribution of data according to a third variable.	Bubble Like a scatterplot, but with bubbles representing categories.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Used with Line - Good for showing the change of data over time. Don't forget to emphasize the area.	Dot plot See above. Don't forget the dots are not always vertical. They can also be horizontal.	Dot map Used for hierarchical visualization. Don't forget to emphasize the hierarchy.	Geographic bubble map For drawing unstructured data. Don't forget to emphasize the geographic bubble map.	Network Used for showing the connections between entities. Don't forget to emphasize the network.
CF heatmap A good way of showing the relationship between two variables. Don't forget to emphasize the correlation.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Used with Line - Good for showing the change of data over time. Don't forget to emphasize the area.	Dot plot See above. Don't forget the dots are not always vertical. They can also be horizontal.	Dot map Used for hierarchical visualization. Don't forget to emphasize the hierarchy.	Dot map For drawing unstructured data. Don't forget to emphasize the dot map.	Cloud A circular field diagram showing the flow of information. Don't forget to emphasize the cloud.
Dot plot A good way of showing the relationship between two variables. Don't forget to emphasize the correlation.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Used with Line - Good for showing the change of data over time. Don't forget to emphasize the area.	Dot plot See above. Don't forget the dots are not always vertical. They can also be horizontal.	Dot map Used for hierarchical visualization. Don't forget to emphasize the hierarchy.	Dot map For drawing unstructured data. Don't forget to emphasize the dot map.	Network Used for showing the connections between entities. Don't forget to emphasize the network.
Dot plot A good way of showing the relationship between two variables. Don't forget to emphasize the correlation.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Used with Line - Good for showing the change of data over time. Don't forget to emphasize the area.	Dot plot See above. Don't forget the dots are not always vertical. They can also be horizontal.	Dot map Used for hierarchical visualization. Don't forget to emphasize the hierarchy.	Dot map For drawing unstructured data. Don't forget to emphasize the dot map.	Cloud A circular field diagram showing the flow of information. Don't forget to emphasize the cloud.
Dot plot A good way of showing the relationship between two variables. Don't forget to emphasize the correlation.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Good for showing individual points in a large dataset. Don't forget to emphasize the distribution.	Dot plot Used with Line - Good for showing the change of data over time. Don't forget to emphasize the area.	Dot plot See above. Don't forget the dots are not always vertical. They can also be horizontal.	Dot map Used for hierarchical visualization. Don't forget to emphasize the hierarchy.	Dot map For drawing unstructured data. Don't forget to emphasize the dot map.	Network Used for showing the connections between entities. Don't forget to emphasize the network.

Visual vocabulary

<https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf>

Human Perception

Vision

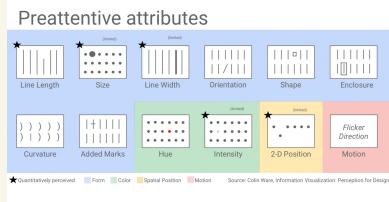
70%
Two Tone

Attention has limits - vision too
We only consume minimal amounts of data, and the brain fills in the rest - our brain's interpretation of **context**
of our sensory neurons are **vision** make data into sound!

How perception works

1. parallel processing
marks each datapoint
channels different encodings (color, shape)
effectiveness: Accuracy, Pop-out, Discriminability, Separability &

Perceptual Hierarchy



Categorical (Nominal)	Categorical (Ordinal)	Quantitative (All)
Position Color Hue Texture Connection Containment Density Color Saturation Shape Length Angle Slope Area Volume	Position Density Color Saturation Color Hue Texture Connection Containment Length Angle Slope Area Volume Shape	Position Length Angle Slope Area Volume Density Color Saturation Color Hue Texture Connection Containment Shape

Always! depends on the data

Gestalt Principles



How humans group things together
Proximity things close together are groups
Similarity things of similar color/shape are groups
Segregation things are enclosed in common groups
Connectivity things are connected to display trends
Continuity things continue to follow smooth path
Closure gaps assume smooth path as well

Factfulness

Best Practices in information design..

General Principles

(1) Do we need the visual? For a look-up tables might be better ☺

(2) What is the insight? Make the insight obvious!

Data Ink: Enhance the data, diminish/minimize distractions

Avoid redundant information taking up space!

Axes: Axes should be placed closely to the highlight!

Aspect Ratio: Change width & height of visuals.

Orientation: we have limited horizontal space; infinite vertical

- Adjacent colors create harmony, opposites create contrast!
- Colors are perceived and associated differently by individuals
- Use more natural softer tone color, with less saturation; start soft, end with saturation and emphasis
- Don't use color just because you can " "
- Beware of colour blindness → orange - blue



Rules of Thumbs ☺



7 Pie Slices



8 Colors

Recommen dated Maximums!

Problematic Charts

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie Slices



8 Colors

7 Lines

10 Bars

7 Pie S

Network Visualization

What does a good network visual look like?

Classic building blocks

Nodes & Edges! BUT! we often get hairballs.

Slides; cytoscape 3.8.2 tools... but can be done in graphviz

Node Concepts

Node size: unimportant $\leftarrow \circ \circ \circ \circ \circ \circ \rightarrow$ important

f.ex. degree, betweenness centrality etc.

REMEMBER that the scale does not need to be linear!

Node color: $\bullet \bullet \bullet \bullet \bullet \bullet \leftarrow$ categories!

REMEMBER colors should be easy to distinct... colorbrewer
find a color-blind friendly palette!

Node labels: context! BUT! please avoid...

Edge Concepts

Edge Thickness: unimportant $\leftarrow | | | | | | \rightarrow$ important

★ use to contrast/complement node size

Edge Color: $| | | |$ \rightarrow should be avoided

gradient color might be beneficial

Edge transparency: With transparent edges, dense areas automatically appear!

Edge labels: DON'T!

Network Layouts

see slides for examples

We have more freedom of movement than usual!

Force-Directed: nodes repel each other, BUT connected nodes attract each other as well

t-SNE: removes some edges... i guess?

Circular: kind of strange... how do we decide order of nodes?

Edge bundling/bending is helpful!

Design Thinking

Think of data viz like a UX designer!
Recommendation: The design of everyday things

Mantra

Overview

Zoom and Filter

Details on-demand

User Tasks

usually action and target

Discover distribution, Compare trends

Iteration

We build something, test it, evaluate ... repeat!

Dieter Principles

Dieter Rams 10 principles of good design
about making design intuitive to use!

Normans Principles

Affordances usage purpose is obvious

Conceptual

Constraints Physical, Semantic & Cultural

Feedback

Aesthetics

What is beauty?

- dynamic something that moves unexpectedly
- harmonious the harmony between items
- Inherent contradicts expectation

Visualize sex worker statistics in DK

Geospatial Visualisation

Why maps?

gives good intuition of spatial differences!

An abstraction

Maps are always abstractions... We can never depict it exactly 1:1, and we always make a choice on what is important! All maps lie! and we should always consider which projection we want to use!

Mercator good for navigation, but makes sizes far from the equator way too big!

Robinsons sacrifices a bit of shape/area but gains balance for the eye!

And more...

Data Ink

is important to consider → do not let the map overpower the data!

Terminology

Storytelling

Why?

Persuasion

We want to elevate our data visualization to a story that will be remembered!

via **Rhetoric** exact truth-telling → supported arguments
via **story** unlocking emotions and discussion

Storytelling is used in many places!

We want to resonate both with logic and emotion

3 step tactic

1)

2)

3)

Journalism

Data Wrapper, Raw Graphs & Flourish

Editorial Thinking! How data becomes a story...

Angles → find core curiosity, and ask questions from many different angles, to satisfy curiosity

Faming → Make choices about what to keep and what to exclude! zoom, focus, perspective!

Focus → Focus on what is important to your story!

The annotation layer is important for understanding the context

Audience People

Don't be your own audience!

Think about the CEO that only quickly views the chart... And the employee that wants to deep dive in data...

Displaying Uncertainty

Sources

We can get uncertainty from many places; sampling, modeling, displaying

Designing nothing

we often have null-values and blanks..
by just ignoring these, we actually ignore
potentially interesting insights!

But how do we describe uncertainty?

Be aware of how people interpret numbers
and probability / fractions in context...



Error Margin

Display a confidence interval, but
beware of the interpretation of the
visual encoding... how do we make
sure the message is clear? and that
the uncertainty is understood?

We're going to The Dark Side

Bias

We have different biases...
f.ex. continuity bias about graphs

Illusory correlations

We might unintentionally deceive by
not displaying all relevant data...
why might there be less cancer;
↳ undiagnosed
↳ lifestyle (pollution etc.)

Our brains want to find patterns
(even if there are none...)

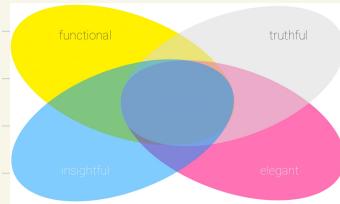
Objectivity

... does not exist! There are always
choices and decisions.
We want to try to be honest!

End-of-course

Numbers tell a story \rightarrow we give it a voice!
But! how we visualize is **so** important!

What have we learned?



Area is important \rightarrow watch out for truncating axes... can be misleading to the human brain!

Data Ink \rightarrow non-data ink should be minimal, as it creates a lot of clutter distracting the eyes

Distinct colors \rightarrow it's important to make colors distinct to easily distinguish different series!

Pie charts \rightarrow throw them out!

Comparing angles is hard \hookrightarrow

The purpose of visuals is to gain insights... don't make it hard!

Good Analysts
Pain Points

are humble storytellers!

Issues still exist!

Time series can be visualized many different ways

Forecasting shows uncertainty \rightarrow more uncertain over time

Outliers remove? footnote? keep? it is the story!
use different axis..