C S 5 3 4 6

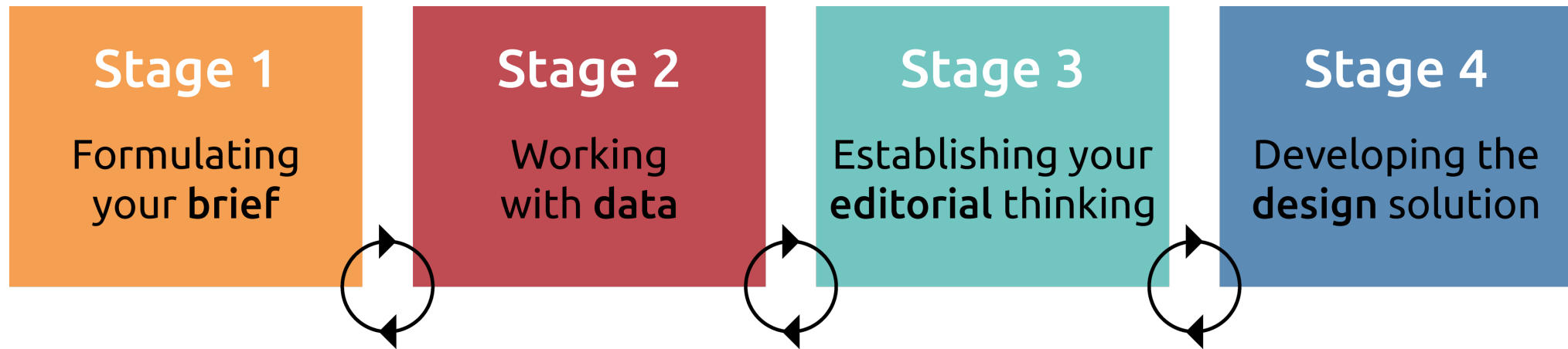# I N F O R M A T I O N   V I S U A L I Z A T I O N

THE DESIGN PROCESS – First 3 Stages

Ref: Chapters 3-6, Data Visualization by Andy Kirk
Ref: Chapter 2, Visualization Analysis and Design by Tamara Munzner

Assoc Prof. Bimlesh Wadhwa | School of Computing | NUS | bimlesh@nus.edu.sg

Visualization Design Process

A 4 Step process

Stage 1	Stage 2	Stage 3	Stage 4
Formulating your **brief**	Working with **data**	Establishing your **editorial** thinking	Developing the **design** solution

Note: There are good and bad solutions. There are no perfect visualizations.

Stage 1: Preparatory stage

Planning, Defining and initiating your project

What is a Project Brief?

- Establish the Goals and Context for your project

Why you are producing this data visualisation –What is the **core curiosity** ?

- Personal intrigue – 'I wonder what …'
- Stakeholder intrigue – 'He/she needs to know …'
- Audience intrigue – 'They will need to know …'
- Anticipated intrigue – 'They might be interested in knowing …'
- Potential intrigue – 'There might be something interesting …'

Example

Project - 'Filmographics' – about the careers of different movie stars.

Reasoning for taking up the project :
I have a passion for movies.
I have a particular curiosity about the emergence, re-emergence and apparent disappearance of certain actors.

The **core curiosity** that triggered this project:

What is the pattern of success or failure in the movie careers of a selection of notable actors?

Note: Quite often one's initial expression of curiosity shifts considerably once data has been gathered and analysed.

Example

Suppose you work for a movie studio

Task given by **a casting director** : compile a one-off report that will profile which actors are potentially the best option to cast in a major sci-fi movie that has just been given the green light to begin production.

Criteria :

they have to be female,

aged 30-45,

must fit the description of 'rising' star,

must not have been in other sci-fi movies,

must not have 'baggage' associated with huge flops,

fees < $2 million.

The **core curiosity** that triggered the project:

What is the career paths of some of the most likely stars who fit the above criteria?

Example

Suppose you work for a business involved in the analysis and review of the state of the movie industry. Your company specialises in producing information for a broad group of users
    - Film industry executives, studio senior management and casting agents, etc.

These **users** have professional curiosities about the movie career statuses.

Curiosity is shaped out of –
    knowing **what *will* be needed** and
    reasonably anticipating **what *could* be needed**.

The **core curiosity** that triggered the project:
What are the different aspects of the industry, covering current trends and the career fortunes of a wide range different actors, that can help the users to identify who is hot, who is not, who is emerging, who is declining, who will cost what, who scores well with different audiences, etc.

Note: Not all the information provided will be of interest all the time to all who consume it but it is provided and available as and when they do need it.

Example

Suppose you work in the graphics editing team for a newspaper.

Public-attention : relatively late-career breakthrough of a certain actor, from TV to movie blockbusters.

Chief editor assigns : examine the fortunes of this actor's career alongside a selection of other actors to provide contrast or draw comparisons with.

The **Core curiosity** is likely driven by the current affairs:

What is newsworthy and likely to be of some interest to readers? The belief is that the analysis of this actor's career path compared with others is aligned to the current entertainment news agenda.

Example

You are interested in data visualisation.
You are also interested in movies
You discover a clean dataset full of rich content about movies and actors.

No specific **core curiosity** as an entry point, except that data offers potential insights.

To fully understand qualities and potential data offers and generally research the movie industry and actors' career paths.
To have a better idea of a more specific curiosity you might pursue.

Summarizing – for your course project

Give your project a Title

Provide clarity about the curiosity that you might be pursuing.
- even in the smallest and seemingly simplest dataset, there are many possibilities for conducting different types of analysis.

Challenge : to know in advance what might be the most interesting and relevant angle to focus on.

Who are the stakeholders and what is their interest?
Who are the intended recipients – the audience – and what is their interest?

The involvement of **people** create influential forces (good and bad) shaping your visualisation thinking.

Summarizing – for your course project

- Stakeholders:
  - request/commission you to do a visualisation

- Audience:
- viewers or consumers of your visualizations

- try to understand their relationship with the subject matter.
- how informed are they and what motivation do they have
  e.g. a passing interest or a definable need?
- their graphical/visualization literacy
- their personal traits can influence your design choices

Summarizing – for your course project

Attempt to elaborate on your Project's Purpose beyond just stating the core curiosity.

What are you trying to accomplish with this visualisation? What is a successful 'outcome'?

Certain types of visualisation will offer a better fit for your project.
- Explanatory
- Confirmatory
- Exploratory

Stage 2 : Preparatory stage

Stage 2

Working
with **data**

Data gathering, handling and preparing your data

Working with Data
some basic concepts and terminology

Two categories of Data Assets – Datasets and Isolated Values

- Datasets : data that exist in tables.

- Isolated Values : data that exists as individual facts and figures.

Tabulated datasets are what we are mainly interested in this course.

Two types of Tabulated Datasets – Normalised and Cross-Tabulated

CATEGORY	MOVIE TITLE	CRITIC RATING	REVIEW GROUP
Star Wars	Star Wars: Episode IV - A New Hope	8.3	Fresh
Star Wars	Star Wars: Episode V - The Empire Strikes Back	8.7	Fresh
Star Wars	Star Wars: Episode VI - Return of the Jedi	6.8	Fresh
Star Wars	Star Wars: Episode I - The Phantom Menace	5.8	Rotten
Star Wars	Star Wars: Episode II - Attack of the Clones 3D	6.6	Fresh
Star Wars	Star Wars: Episode III - Revenge of the Sith 3D	7.2	Fresh
X-Men	X-Men	7.0	Fresh
X-Men	X2: X-Men United	7.4	Fresh
X-Men	X-Men: The Last Stand	5.9	Rotten
X-Men	X-Men Origins - Wolverine	5.1	Rotten
X-Men	X-Men: First Class	7.4	Fresh
X-Men	The Wolverine	6.3	Fresh
X-Men	X-Men: Days of Future Past	7.6	Fresh
Tolkien	The Lord of the Rings: The Fellowship of the Ring	8.2	Fresh
Tolkien	The Lord of the Rings: The Two Towers	8.5	Fresh
Tolkien	The Lord of the Rings: The Return of the King	8.7	Fresh
Tolkien	The Hobbit: An Unexpected Journey	6.6	Fresh
Tolkien	The Hobbit: The Desolation of Smaug	6.8	Fresh
Tolkien	The Hobbit: The Battle of the Five Armies	6.3	Fresh

CATEGORY	MOVIES	MAX CRITIC RATING	REVIEW: FRESH	REVIEW: ROTTEN
Star Wars	6	8.7	5	1
Tolkien	6	8.7	6	0
X-Men	7	7.6	5	2
SUMMARY	19	8.7	16	3

Normalised dataset

Looks like lists of data values.

Typically as a series of columns and rows of data, Or an arrangement of fields and records.

CATEGORY	MOVIE TITLE	CRITIC RATING	REVIEW GROUP
Star Wars	Star Wars: Episode IV - A New Hope	8.3	Fresh
Star Wars	Star Wars: Episode V - The Empire Strikes Back	8.7	Fresh
Star Wars	Star Wars: Episode VI - Return of the Jedi	6.8	Fresh
Star Wars	Star Wars: Episode I - The Phantom Menace	5.8	Rotten
Star Wars	Star Wars: Episode II - Attack of the Clones 3D	6.6	Fresh
Star Wars	Star Wars: Episode III - Revenge of the Sith 3D	7.2	Fresh
X-Men	X-Men	7.0	Fresh
X-Men	X2: X-Men United	7.4	Fresh
X-Men	X-Men: The Last Stand	5.9	Rotten
X-Men	X-Men Origins - Wolverine	5.1	Rotten
X-Men	X-Men: First Class	7.4	Fresh
X-Men	The Wolverine	6.3	Fresh
X-Men	X-Men: Days of Future Past	7.6	Fresh
Tolkien	The Lord of the Rings: The Fellowship of the Ring	8.2	Fresh
Tolkien	The Lord of the Rings: The Two Towers	8.5	Fresh
Tolkien	The Lord of the Rings: The Return of the King	8.7	Fresh
Tolkien	The Hobbit: An Unexpected Journey	6.6	Fresh
Tolkien	The Hobbit: The Desolation of Smaug	6.8	Fresh
Tolkien	The Hobbit: The Battle of the Five Armies	6.3	Fresh

Example : normalised data where the **columns of variables** provide different **descriptive values** for **each movie (or record)** held in the table.

17

Cross-Tabulated Dataset

- aggregated calculations instead of raw data values

e.g. the results of statistical operations

(like summed totals, maximums, averages)

CATEGORY	MOVIES	MAX CRITIC RATING	REVIEW: FRESH	REVIEW: ROTTEN
Star Wars	6	8.7	5	1
Tolkien	6	8.7	6	0
X-Men	7	7.6	5	2
SUMMARY	19	8.7	16	3

Example : cross-tabulated result of the normalised table of movie data

A statistical summary for each movie category.

'Max Critic Rating' is formed from an aggregating calculation based on the 'Critic Rating' and 'Category' in the normalised dataset

Two types of Data – Qualitative and Quantitative

Nominal (Qualitative)	**Ordinal (Qualitative)**
Interval (Quantitative)	**Ratio (Quantitative)**

NOIR
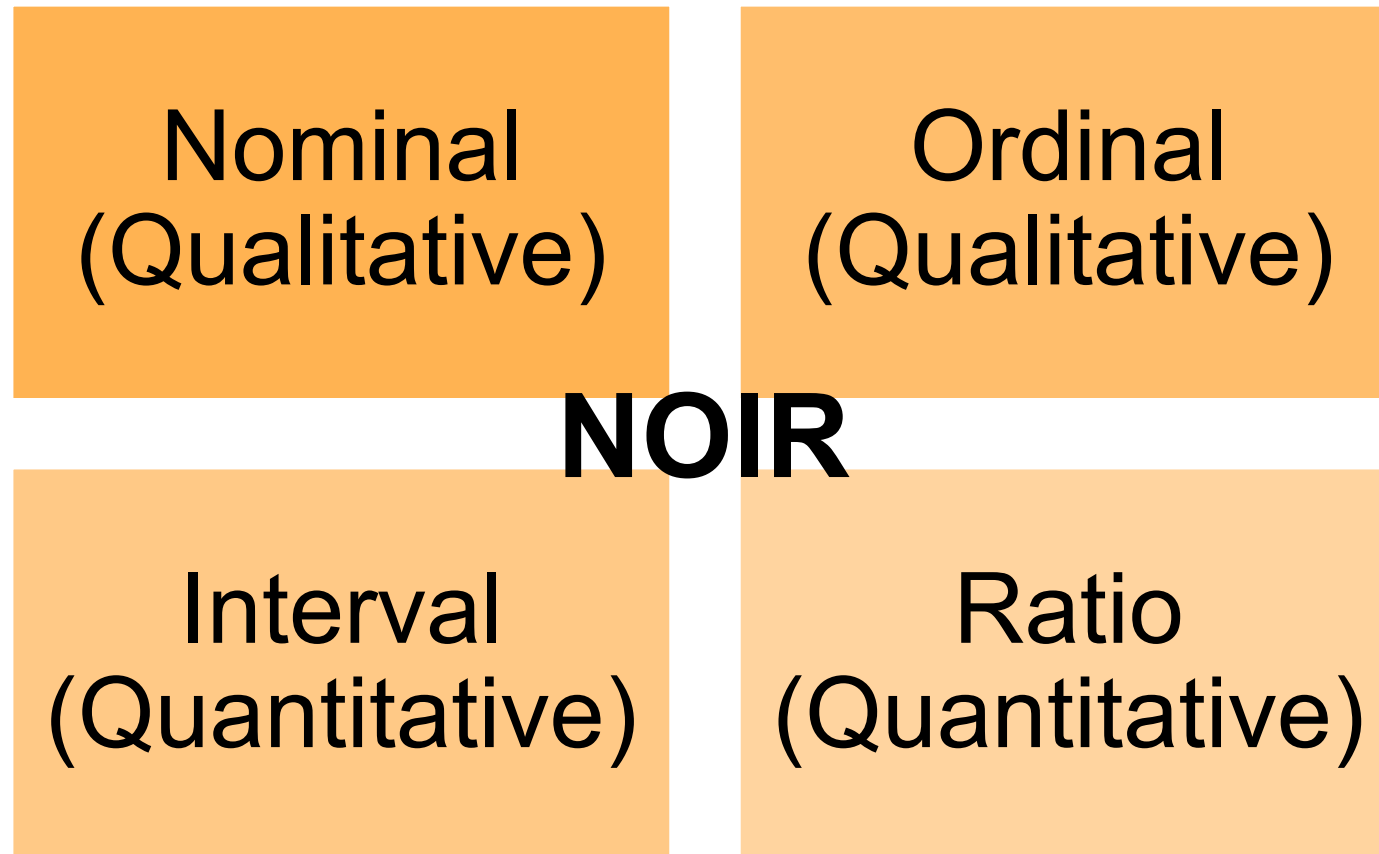
Nominal (Qualitative)

It exists in **categorical** form

Categories - a means of distinguishing, labelling and organising values.

Examples:
Location names
University Department names
Languages
Athletic events

Student ID number
Jersey number
TRUE or FALSE or a 1 and a 0
1 (male), 2(female) and 3 (other)

Nominal (Qualitative)

There might be **a hierarchical relationship** between two or more nominal variables
   - major category and sub-category

   Examples

   Major category: Country' and a sub-category : 'Airport';
   Major category: Industry' and a sub-category: Company
   Names

Ordinal (Qualitative)

- **Categorical** in nature but with **characteristics of order**.
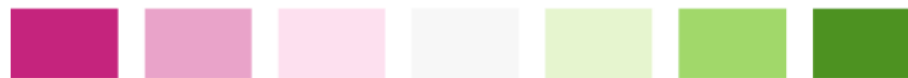
Examples:
- Scale values of 1 (unhappy) to 5 (very happy).
- Weather forecast expressed as Very Hot, Hot, Mild, Cold, Freezing.
- Academic rank of a conference.
- Delivery options for an Amazon order: Express(Today), Next Day, Standard(2-5 days).
- Medals categories for an athletic event: Gold, Silver, Bronze.

Ordinal (Qualitative)

Sometimes we find a a natural 'pivot' where **the direction of order** changes around a recognisable mid-point.

Example: happiness scale   …, …., unhappy, 'no feeling', feeling good, …. , ….

Interval (Quantitative)

- Numeric measurements
  - Each value at an **equal distance** (interval) from another value.
  - They **can be added or subtracted** but **cannot be multiplied or divided**.
  - There is **no true zero** on an interval scale.

- Examples

  Test scores
  e.g. SAT scores range from 200 to 800
  There is no absolute zero because there are no scores below 200.

  Time
  E.g. Using a twelve-hour clock.
  There is no absolute zero
  Difference between 4 and 5 o'clock is the same as the difference between 5 and 6 o'clock.

24

Ratio (Quantitative)

- Numeric measurements
  has a **defined zero point**

Examples :

Income

Height

Weight

Annual sales

Age

Rainfall

Budget

A few other types of data :

<table>
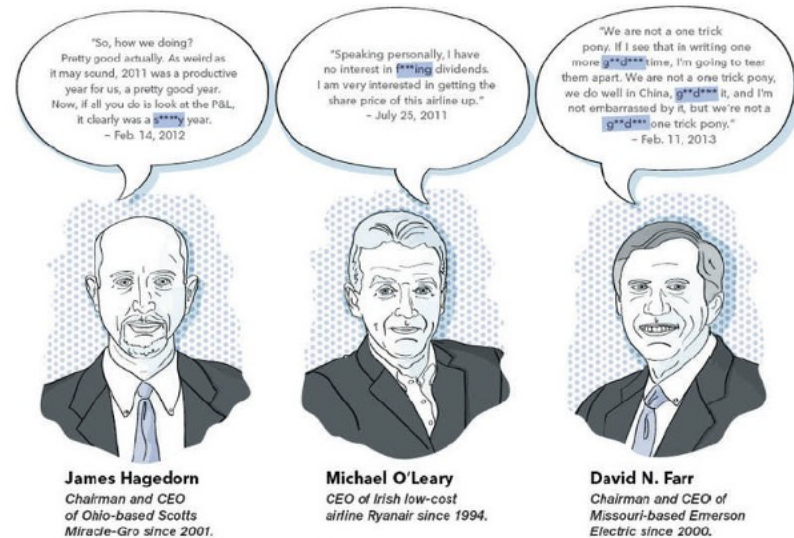<tr><td>Textual Data</td><td>Temporal data</td></tr>
<tr><td>Discrete Data</td><td>Continuous Data</td></tr>
</table>

Textual (Qualitative)

Generally, as unstructured streams of words.
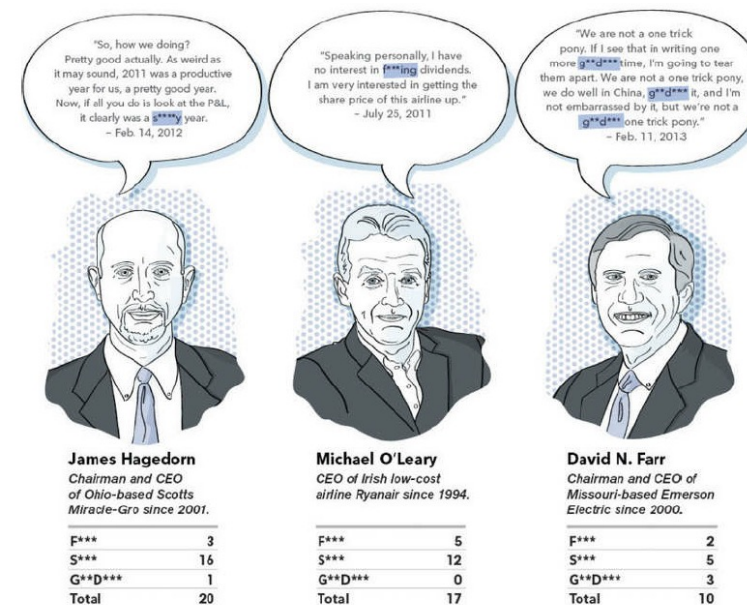
Examples

- Comments
- Descriptions
- Titles
- URL

Textual (Qualitative)

Typically, you can **derive** **categorical classifications** and **quantitative measurements**
- **e.g.** to count the use of certain words compared to others and track their usage over time.

The original form of the textual data can be of direct value for annotation purposes, **e.g.** to include as captions.

Usage by word, 2004 to 2014[1]

197 S***
34 G**D***
17 F***
6 A**H***

"So, how we doing? Pretty good actually. As weird as it may sound, 2011 was a productive year for us, a pretty good year. Now, if all you do is look at the P&L, it clearly was a s****y year." – Feb. 14, 2012

"Speaking personally, I have no interest in f***ing dividends. I am very interested in getting the share price of this airline up." – July 25, 2011

"We are not a one trick pony. If I see that in writing one more g**d***time, I'm going to tear them apart. We are not a one trick pony, we do well in China, g**d*** it, and I'm not embarrassed by it, but we're not a g**d*** one trick pony." – Feb. 11, 2013

James Hagedorn
Chairman and CEO of Ohio-based Scotts Miracle-Gro since 2001.

F***	3
S***	16
G**D***	1
Total	20

Michael O'Leary
CEO of Irish low-cost airline Ryanair since 1994.

F***	5
S***	12
G**D***	0
Total	17

David N. Farr
Chairman and CEO of Missouri-based Emerson Electric since 2000.

F***	2
S***	5
G**D***	3
Total	10

28

Temporal Data

Different components of time can be positioned against almost all data types

- Ordinal: 'PM', 'Afternoon', 'March', 'Q1'
- Interval: '12', '12/03/2016', '2016'
- Ratio: '16:00'
- Textual: 'Four o'clock in the afternoon on Monday, 12 March 2016'

Discrete Data

- All qualitative data types and any quantitative types for which only a whole is possible

  - variables that have no 'in-between' state.

Examples
- Heads or tails for a coin toss
- Days of the week
- The size of shoes
- Numbers of seats in a theatre

Continuous Data

- Continuous variables can hold the value of an in-between state

    - in theory, any value between the natural upper and lower limits if it is possible to take measurements in fine degrees of detail

Examples:
- Height
- Weight
- Temperature
- Time

Working with data

Data gathering, handling and preparing your data

Working with data

- Data acquisition: Gathering the raw material.

- Data examination: Identifying physical properties and meaning.

- Data transformation: Enhancing your data through modification and consolidation.

- *Data exploration: Using exploratory analysis and research techniques to learn.*

Data Acquisition

- What data do you need?
- From where, and how will the data be acquired?
- When can you obtain it?

Data Acquisition

do you need?

Example

Core curiosity :

'I wonder what a map of McDonald's restaurant openings looks like over time?'.

Dataset : details of all the McDonald's restaurants that have ever opened.

Data items : date of opening, the location details (as specific as possible), Status or a closing date

Get an **initial sense** of your potential data.
- of the timeframe (number of years)
- scale (outlets, global spread)
- differences between what is considered a restaurant
and a franchise positioned in shopping malls or at Transport Hubs
- Will this type of information be available on McDonald's own site?
- Do you have to get in touch with somebody (a human) in the commercial or PR department ?
- Possibly some fast-food enthusiast has already gathered and made available data like this?

Data Acquisition   <mark>What data</mark> do you need?

What if you find a dataset that includes not just
McDonald's restaurants
but ALL fast-food outlets.

→ potentially broadens the scope analysis about the growth of
the fastfood industry at large to contextualise MacDonald's
contribution to this.

Discuss the merits of this wider perspective with the stakeholders.

Data Acquisition   ==What data== do you need?

Think about the resolution of the data you need

- Full resolution - all available data.
- Partial resolution – a sub-set of data

What if you find a BIG dataset

Do you need full resolution ?
– down to the most detailed grain (e.g. all details about all MacDonald's restaurants, not just totals per city or country).

OR a sub-set of your data.
how will you select or exclude records?
- what will be sufficient to be representative of the range and diversity
e.g. you only need a statistical, high-level summary -- total number of restaurants opened by year?

A sensible attitude : 'what I have is good enough to start with' & I shall revisit and collect more data as needed.

Data Acquisition

Sourcing - from where, and how will the data be acquired?

- Primary data collection: measuring and collecting the raw data yourself

- Data foraging: pull data from multiple sources to create a single dataset ; example- researching historical data from archived newspapers that were only published in print form and not available digitally; extracting from pdf files

- Web scraping : use programs to extract structured and unstructured items of data published in web pages. (check and respect the terms of use before undertaking this)

- Curated by Others : Download from the Web; provided by stakeholders; data marketplaces; API access

Data Acquisition

<mark>When</mark> do you acquire?

- Whenever data is ready and available

- Initial sample – for early work, for piloting
- Complete sample- Ideal situation
  - You do not want to progress too far ahead with your visualisation workflow if there is the real prospect that any further data that emerges could offer new insights or even trigger different, more interesting curiosities.

Data Examination   <mark>Familiarising</mark> yourself with your data

- ## Type : identify what data types you have
  - Define each column or field of data
  - qualitative (text, nominal, ordinal) or quantitative (interval, ratio) and discrete or continuous

- ## Size: Within each column or field, what range of values exist and what are the specific attributes/formats of the values held.
  - For example, what is the lowest and the highest value?
  - In what number format is it presented (i.e. how many decimal points or comma formatted)?
  - How many different values are held?
  - What is the maximum character length or word count?

Data Examination

Condition: identifying any problems - quality and completeness issues.

Things to look out for **e.g.**

- Missing values, records or variables - Are empty cells assumed as being of no value (zero/nothing or n/a, null)?

- Erroneous values – Typos and any value that clearly looks out of place (such as a gender value in the age column)

- Date issues around format (dd/mm/yy or mm/dd/yy)

- Inconsistencies – Capitalisation, units of measurement, value formatting.

- Duplicate records.

- Out of date – Values that might have expired in accuracy

- Uncommon system characters or line breaks or leading or trailing spaces.

Data Examination   ==Familiarising== yourself with your data

- Describe and understand the characteristics of each data variable

Frequency counts: of categorical data

Frequency distribution: of quantitative values ; type and shape of the distribution of values.

Measurements of central tendency : of quantitative values, e.g. mean (the average value); median (the middle value if all quantities were arranged from smallest to largest); mode (the most common value).

Measurements of spread - Maximum, minimum and range: the highest and lowest ; magnitude of spread of values, Percentiles: the value below which x% of values fall; Standard deviation

Data Transformation

- Transform **to clean**:
  - 'find and replace' (or remove)
  - convert data into new states For example, "male" to "M" or to "1" or to "MALE"
  - 'sort' or 'filter', to find, isolate and then modify problem values.
  - Eliminate junk

Example: For McDonald's restaurants dataset

Would you need name of the restaurant manager, details of the opening times or the contact telephone number ?

Data Transformation

Transform **to convert or create**:
  Create new data values out of existing ones.
  Form new calculations, values, groupings etc. that expands the available data.

Example
- extracting the constituent parts of a 'Release Date' field in order to group, and use the data in different ways.

- e.g. the 'Month' and 'Year' fields in order to explore within-year and across-year seasonality.

-creating a 'Full Release Date' formatted version of the date possibly for labeling purposes.

Release Date	Month	Year	Full Release Date
21/06/13	6	2013	Friday, 21 June 2013
26/08/13	8	2013	Monday, 26 August 2013
21/12/12	12	2012	Friday, 21 December 2012
22/12/10	12	2010	Wednesday, 22 December 2010
11/06/09	6	2009	Thursday, 11 June 2009
30/11/10	11	2010	Tuesday, 30 November 2010
29/04/07	4	2007	Sunday, 29 April 2007
14/12/07	12	2007	Friday, 14 December 2007
07/11/06	11	2006	Tuesday, 7 November 2006
20/10/06	10	2006	Friday, 20 October 2006
22/12/06	12	2006	Friday, 22 December 2006
12/11/04	11	2004	Friday, 12 November 2004
20/06/03	6	2003	Friday, 20 June 2003
01/01/98	1	1998	Thursday, 1 January 1998
01/01/03	1	2003	Wednesday, 1 January 2003
29/08/03	8	2003	Friday, 29 August 2003

Data Transformation

Examples

Calculate the duration in days from the pairs of 'start date' and 'end date'

Create a new variable that summarises another
 - IF Age < 18 THEN status = "Child", ELSE status = "Adult".

Create a calculation that standardises quantities
e.g. calculating 'per capita' or 'per unit' values.

Data Transformation

Categorical conversions for Textual data

Identify keywords or summary themes from text and convert these into categorical classifications.

Identify instances of certain cases existing or otherwise
(e.g. X is mentioned in this passage)

Identify the existence of certain relationships
(e.g. A and B are both in the same passage, C is always before D)

Use NLP techniques to determine sentiments, to identify specific word types (nouns, verbs, adjectives) or sentence structures (around clauses and punctuation marks).

With URLs: isolate and extract the different components of website address and sub-folder locations.

Data Transformation

Quantitative conversions for Textual Data
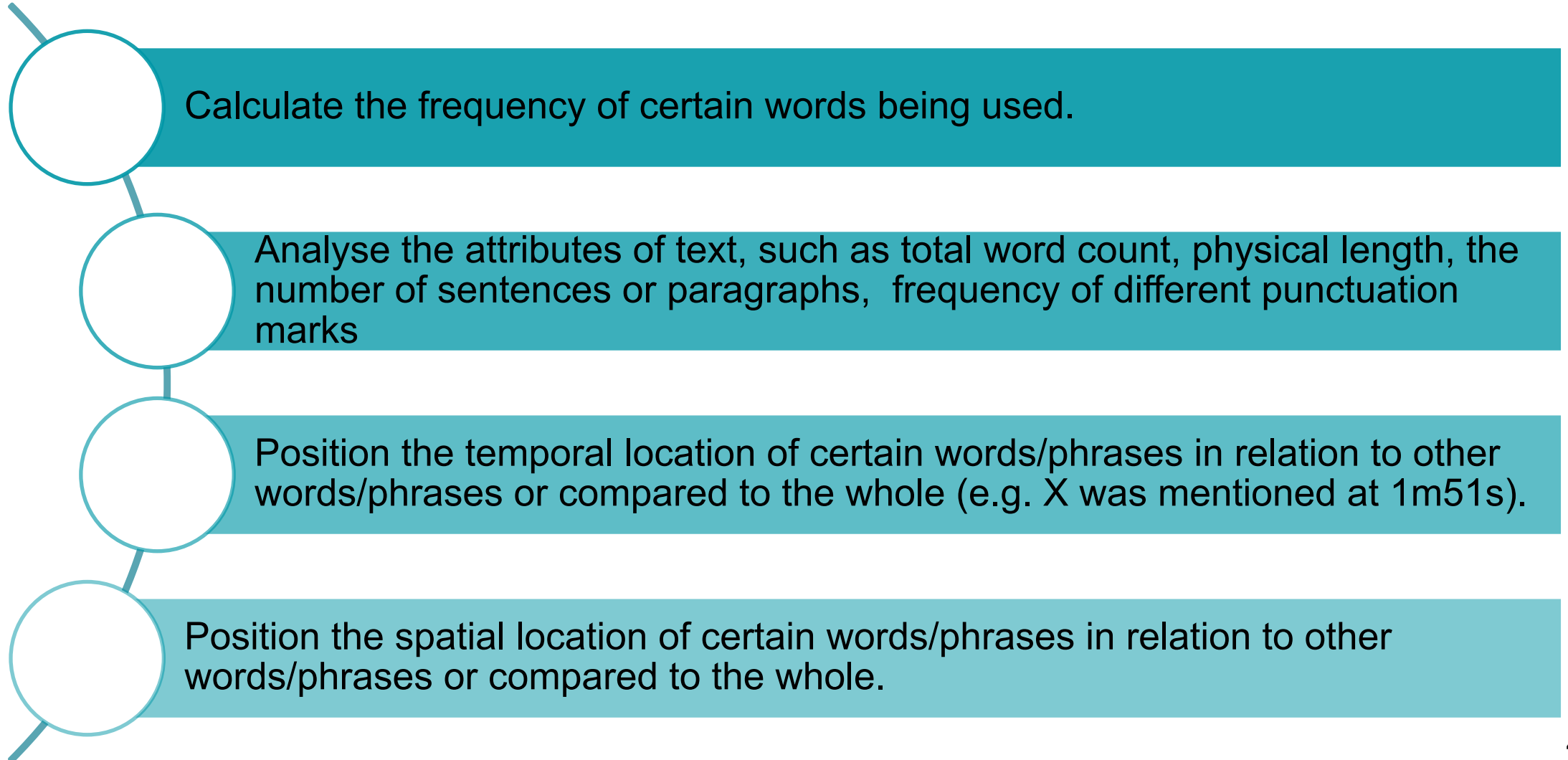
Calculate the frequency of certain words being used.

Analyse the attributes of text, such as total word count, physical length, the number of sentences or paragraphs, frequency of different punctuation marks

Position the temporal location of certain words/phrases in relation to other words/phrases or compared to the whole (e.g. X was mentioned at 1m51s).

Position the spatial location of certain words/phrases in relation to other words/phrases or compared to the whole.

47

Data Transformation

Transform **to consolidate**:

Bringing in additional data - more variables or more records

Example   :

You had details about locations only at country level

You want to group and aggregate at the continent level.

→ Find a dataset that holds values showing the relationships between country and continent

→ Add a new variable to your dataset against which you would perform a simple lookup operation to fill in the associated continent values.

Data Transformation

Transform **to consolidate**:

Bringing in additional data - more variables or more records

Example   :

Making a dataset uptodate by adding newer data

Music record sales up to a certain point in time

but more data becomes available during the course of the project

Data Transformation

Transform **to consolidate**:

Bringing in additional data - more variables or more records

Example   :

Add other media assets to enhance your presentation

e.g. gathering photos (headshots of the people in your data),

icons/symbols (country flags),

links to articles (URLs), or videos (clips of goals scored)

All of these would contribute to broadening the scope of your annotation options.

Note: when making changes, it is useful to preserve original (unaltered) data items nearby for easy rollback should you need them.

Data Exploration

- Be aware of all the options

Data Exploration

The ==*known knowns*==

- knowledge about your subject and about the qualities present in your data that you are aware of
– **you are aware** that **you know** these things.

You cannot afford to be complacent

	ACQUIRED	
	KNOWN	**UNKNOWN**
AWARENESS **KNOWN**	The things we are aware of knowing *Beware complacency*	The things we are aware of not knowing *Deductive reasoning*
UNKNOWN	The things we are unaware of knowing *Acknowledge & retrieve*	The things we are unaware of not knowing *Inductive reasoning*

Data Exploration

The ==*unknown knowns*==
- things you did not realise you knew or

- perhaps did not wish to acknowledge that you knew about a subject.

forgotten, consciously ignored or buried.

You need to acknowledge & Retrieve these.

	ACQUIRED	
	KNOWN	**UNKNOWN**
AWARENESS — KNOWN	The things we are aware of knowing *(Beware complacency)*	The things we are aware of not knowing *(Deductive reasoning)*
AWARENESS — UNKNOWN	The things we are unaware of knowing *(Acknowledge & retrieve)*	The things we are unaware of not knowing *(Inductive reasoning)*

Data Exploration

– The *known unknowns* and *unknown unknowns*

the knowledge that has yet to be acquired

      You can plug these gaps

e.g. through Exploratory work - create a
visualisation that further facilitates understanding
of subject and data.

ACQUIRED

	KNOWN	UNKNOWN
KNOWN	The things we are aware of knowing Beware complacency	The things we are aware of not knowing Deductive reasoning
UNKNOWN	The things we are unaware of knowing Acknowledge & retrieve	The things we are unaware of not knowing Inductive reasoning

AWARENESS

54

Data Exploration

Exploratory – Distribution of Marathon finishing times

Results of nearly every major and many minor (full) marathon from around the world.

On the surface, **the distribution of finishing times reveals the common bell shape**.

Zoom in closer : **some really interesting threshold patterns** for finishing times on or **just before the three-, four- and five-hour marks.**

→ **The influence of runners setting themselves targets**, often **rounded to the hourly milestones**, appeared to **affect the results achieved**.

Arbitrary goals, like round numbers, can be motivating – just ask 9 million marathoners.

Distribution of marathon finishing times

The small spikes are people making their goals, with not a minute to spare. A finishing time of 3:59 is 1.4 times as likely as one of 4:01.
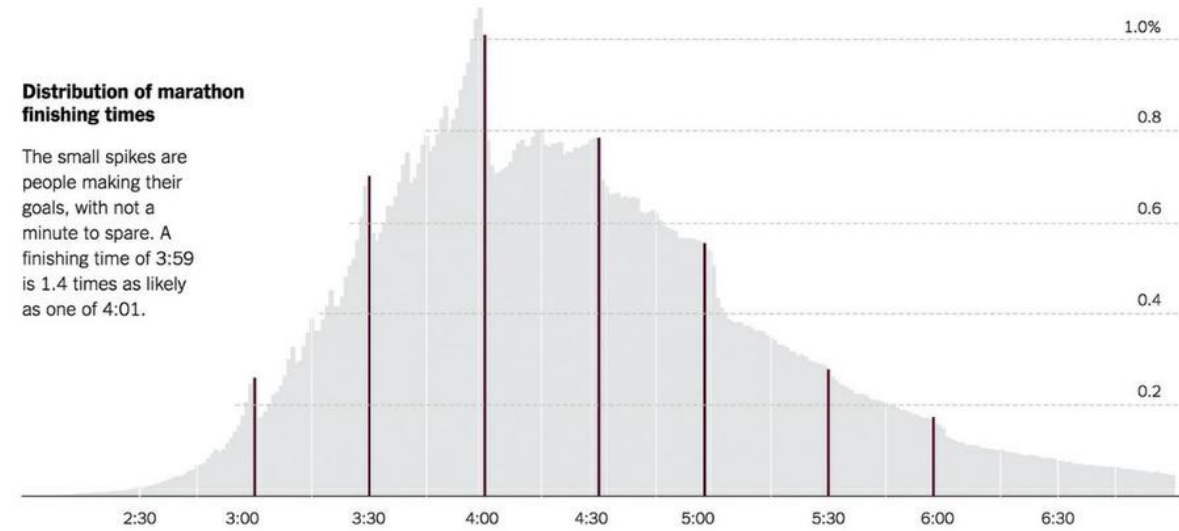
Based on data from Eric Allen, USC, Patricia Dechow, U.C. Berkeley, Devin Pope and George Wu, University of Chicago.

Data Exploration

- Two distinct approaches to plug the gaps
– deductive and inductive

- *Deductive* reasoning (targeted): aware of unknown
    - adopt a detective's mindset
    - you have a specific curiosity or hypothesis, framed by subject knowledge
    - interrogate the data in order to determine whether there is any evidence of relevance or interest in the concluding finding.

- *Inductive* reasoning (open in nature) : unaware of unknown
- 'play around' with the data, based on your sense or instinct about what might be of interest, and see what emerges.

Summarizing – Stage 2: Working with data

- Distinction between normalised and cross-tabulated datasets

- Different types of data NOIR , Text, Temporal, Discrete, Continuous

Summarizing – Stage 2: Working with data

4 steps involved in working with data:

(i) Acquisition : sources and methods for getting your data.
Curated by you: primary data collection, data foraging, web scraping
Curated by others: given to you, downloaded from the Web, third-party
services, APIs.

(ii) Examination : appreciation of the Data characteristics
type, size, condition, meaning, completeness.

Summarizing – Stage 2: Working with data

4 steps involved in working with data:

(iii) Transformation Getting it ready for visualisation design

Clean: resolve any data quality issues.

Create and Convert: consider new calculations and conversions.

Consolidate: what other data (to expand or append) or other assets could be sought to enhance your

(iv) Exploration -

Deepen your familiarity with the data

Using techniques to *explore* the data's qualities

Stage 3: Preparatory stage

Stage 3

Establishing your **editorial** thinking

Defining what you will show your target audience

Editorial perspectives - Angle, Framing and Focus

Editorial perspective – Angle

You typically show data from multiple different angles using several charts presented together.

Angle -- Angle of analysis you intend to show
Angle is relevant to the purpose /goal/ project brief.
What does your intended *audience* want or need to know?
What do *you* want your audience to know?

Example-
how product sales have changed over time, OR
how sales look organised by regional hierarchically OR
how they compare on a map and over time?

Editorial perspective - Angle

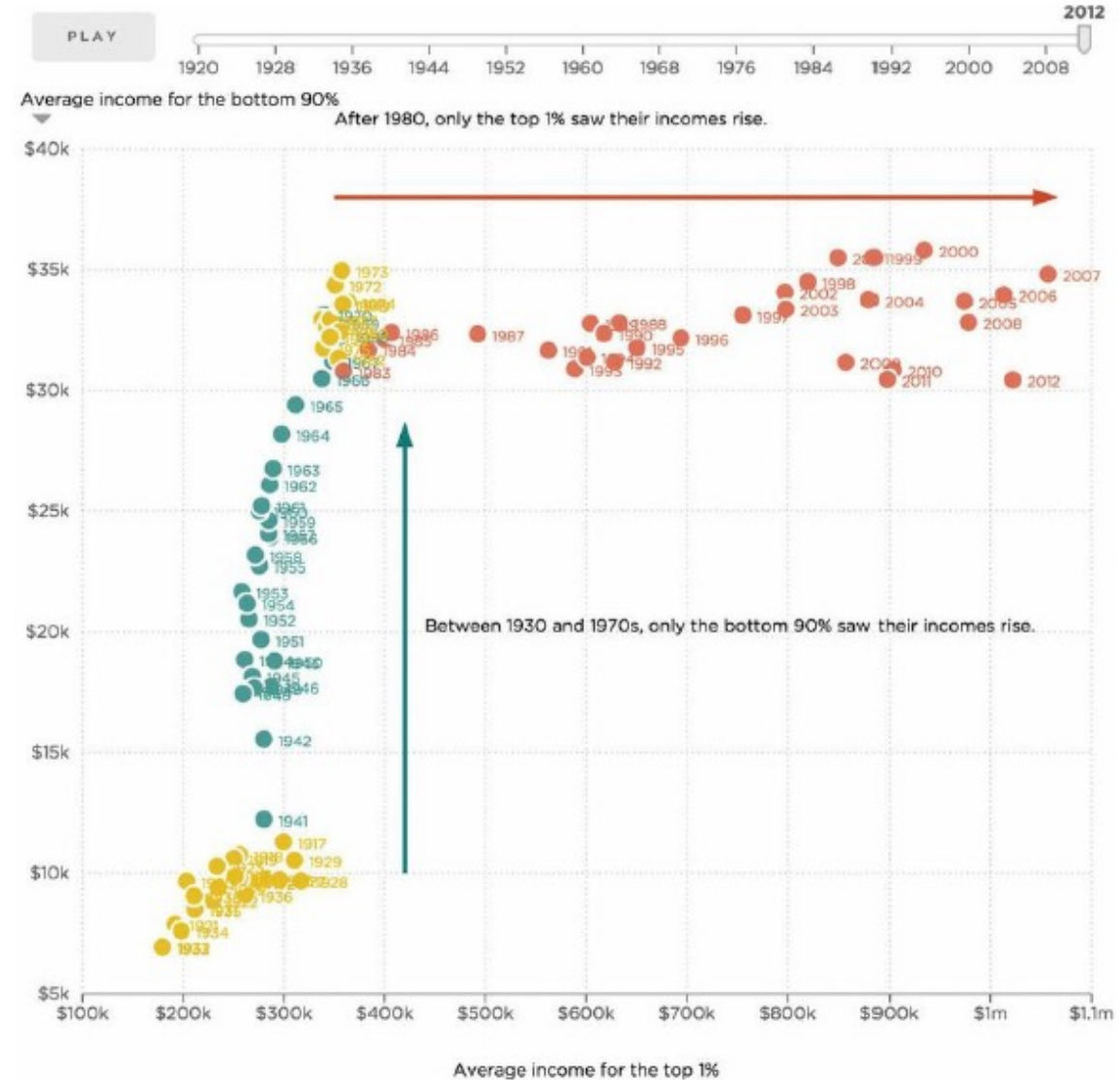Example- Context - Wealth distribution for a country

What is the relationship between two quantitative measures (average income for the bottom 90% and for the top 1% of earners) and how has this changed over time (year)?

This angle is **relevant** because the relationship between the *haves* and the *have-nots* is a key indicator of wealth distribution (the context).

Editorial perspective - Framing

- Framing: what will be included in the scope and what will be left out?

A **filtering decision** concerned with which data to include / exclude:
- All category values, or just a select few?
- All quantitative values or just those over a certain threshold?
- All data or just those between a defined start and end date period?

Example - filters for time period (1917 to 2012) and country (just for the USA).

Editorial perspectives - Focus

Focus : emphasising what is more important in contrast to what is less important.

- This is about reducing noise.
  If everything in a visualisation is shouting, nothing is heard;
  if everything is in the foreground, nothing stands out;
  if everything is large, nothing is dominant.

Example-
Users to focus incrementally through each year :
eg using a **time slider** ; using **colouring** for each consecutive year marker for emphasis.

Example- Angle

What is the relationship between two quantitative measures (average income for the bottom 90% and for the top 1% of earners) and how has this changed over time (year)?

Example – Framing

Filters for time period (1917 to 2012) and country (just for the USA).

Example- Focus

Users to focus incrementally through each year : eg using a **time slider** ; using **colouring** for each consecutive year marker for emphasis.

PLAY

2012

1920 1928 1936 1944 1952 1960 1968 1976 1984 1992 2000 2008

Average income for the bottom 90%

After 1980, only the top 1% saw their incomes rise.

Between 1930 and 1970s, only the bottom 90% saw their incomes rise.

Average income for the top 1%

Editorial perspectives

Influence on Design Choices

(i) The **angle shapes the data representation** approach.
In lay terms, it determines **which chart type** is used.

Example- relationship between two quantitative measures over time (average income for bottom 90% vs. top 1% of earners).
A suitable chart type to portray this visually is the scatter plot (as selected).
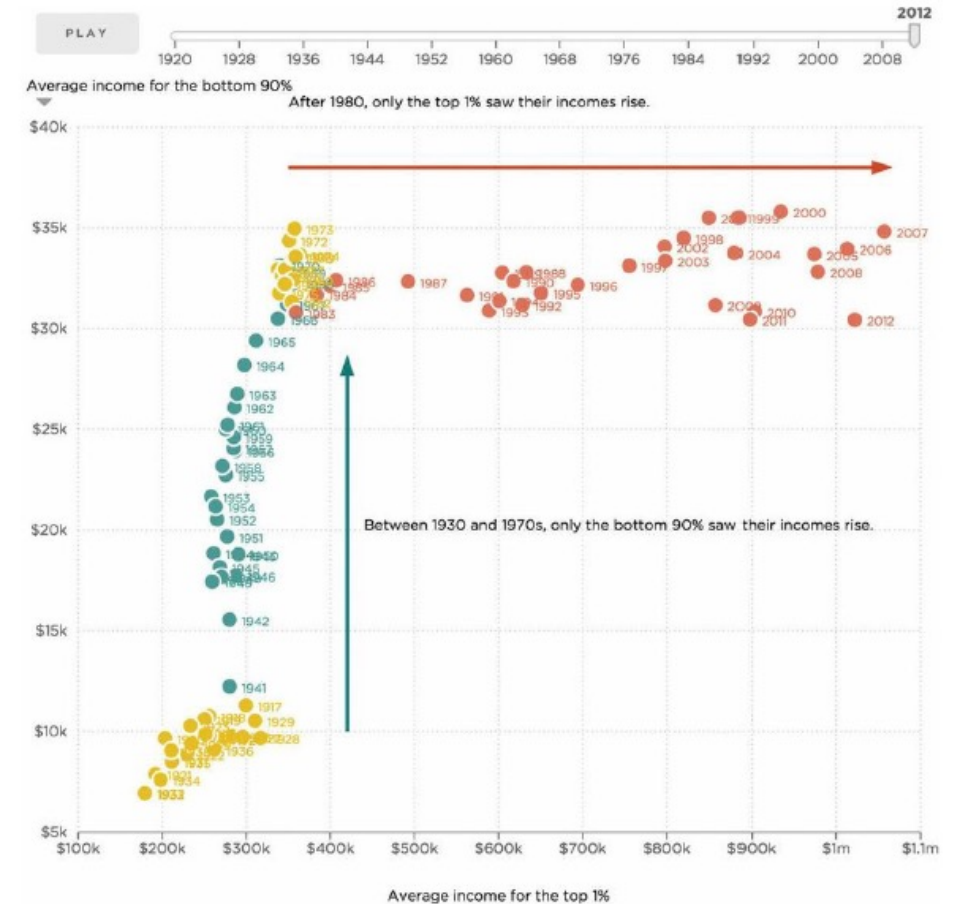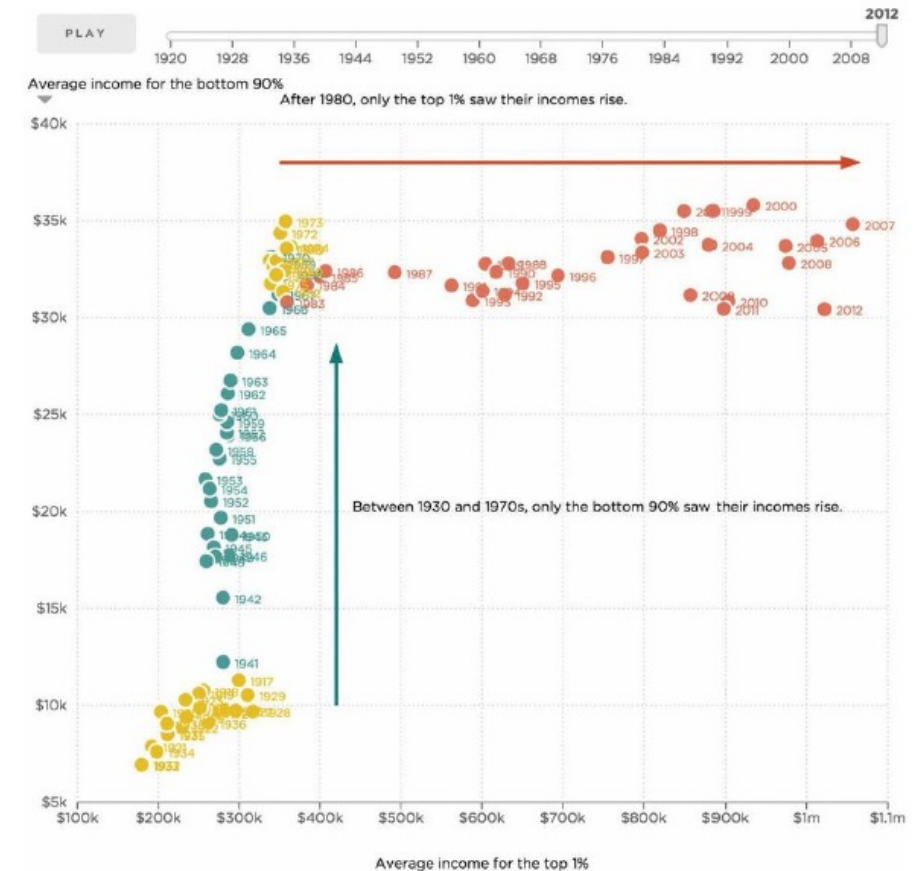
Editorial perspectives   Influence on Design Choices

(ii) Interactivity can be influenced by the editorial decisions concerning **focus**:

- Example - the yearly values sequences & the emphasis on each consecutive value

→ the '**time slider**' control

Editorial perspectives   Influence on Design Choices

(iii) Another example of **Focus** – draw viewers' attention to the two prominent patterns that support the general fall and then rise of inequality.

- **Annotation:** The chart annotations show **two arrows and associated captions**, drawing attention to the two prominent patterns that support the general fall and then rise of inequality.
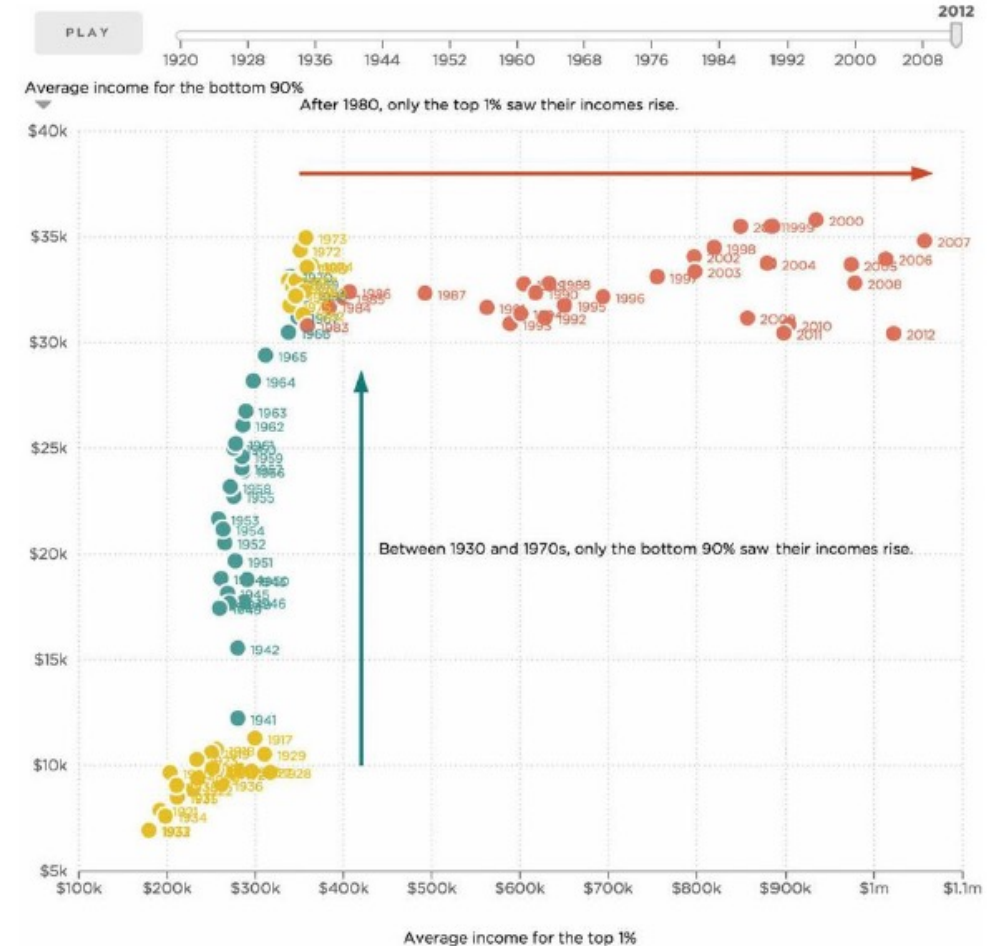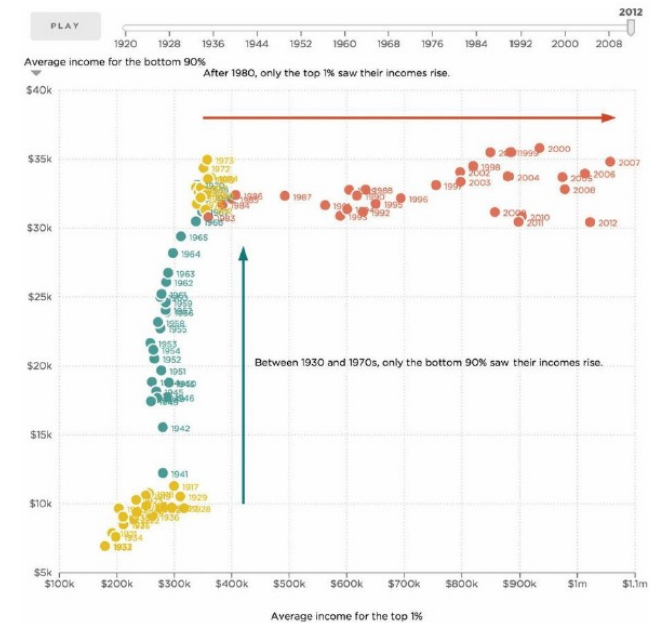
Editorial perspectives   Influence on Design Choices

(iv) Colour: to emphasise content and direct the eye.

Editorial *focus* would influence the decision to deploy colour states within the chart
 - a default colour to show all points at the start of the animation and
- three different emerging colours to separate the three clustered groups visually.

Summarizing – Stage 3 : Editorial Thinking

Angle

- Must be relevant in its potential interest for your audience.

- Must have sufficient quantities to cover all relevant views – but no more than required.

Framing

- Applying filters to your data to determine the inclusion and exclusion criteria

Focus

- Which features of the display to draw particular attention to? How to organise the visibility and hierarchy of the content?