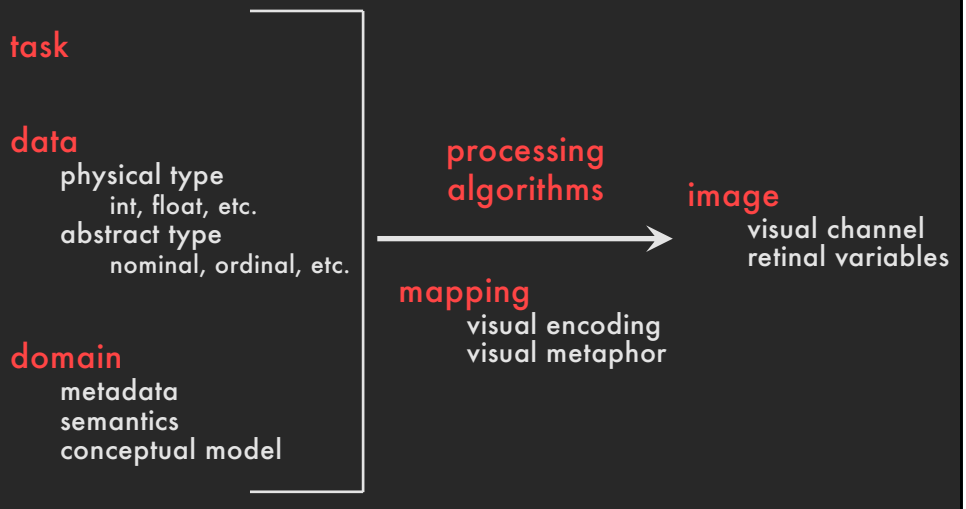
Visualization Design

Maneesh Agrawala

CS 448B: Visualization
Fall 2018

Last Time: Data and Image Models

The big picture



[based on slide from Munzner]

Nominal, ordinal and quantitative

On the theory of scales of measurements
S. S. Stevens, 1946

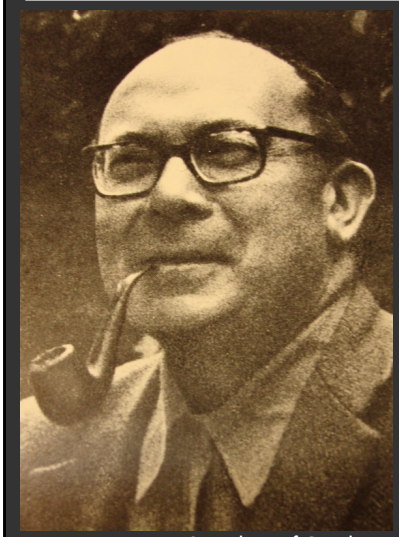
N - Nominal (labels)
Fruits: Apples, oranges, ...
Operations: =, ≠

O - Ordered
Quality of meat: Grade A, AA, AAA
Operations: =, ≠, <, >, ≤, ≥

Q - Interval (location of zero arbitrary
Dates: Jan, 19, 2006; Loc.: (LAT 33.98, LON -118.45)
Like a geometric point. Cannot compare directly
Only differences (i.e. intervals) may be compared
Operations: =, ≠, <, >, ≤, ≥, -

Q - Ratio (location of zero fixed)
Physical measurement: Length, Mass, Temp, ...
Counts and amounts
Like a geometric vector, origin is meaningful
Operations: =, ≠, <, >, ≤, ≥, -, ÷

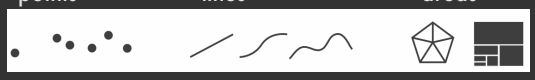
Marks and Visual Variables



Semiology of Graphics
J. Bertin, 1967

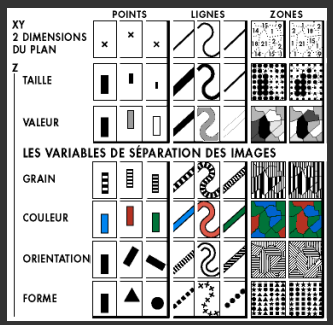
Marks: geometric primitives

points lines areas

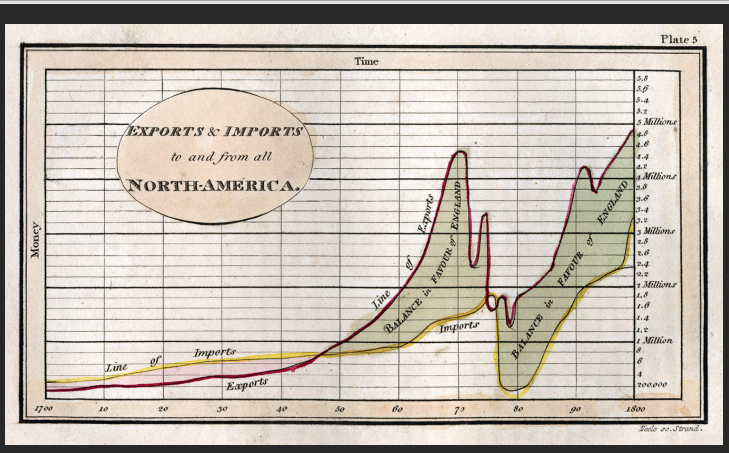


Visual Variables: control mark appearance

- Position (2x)
- Size
- Value
- Texture
- Color
- Orientation
- Shape



Playfair 1786/1801



- Time → x-position (Q, linear)
- Exports/Imports Values → y-position (Q, linear)
- Exports/Imports → color (N, O, nominal)
- Balance for/against → area (maybe length??) (Q, linear)
- Balance for/against → color (N, O, nominal)

Bertins' "Levels of Organization"

Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	O	
Color	N		
Orientation	N		
Shape	N		

N Nominal
O Ordered
Q Quantitative

Note: Q < O < N

Note: Bertin actually breaks visual variables down into differentiating (≠) and associating (≡)

Principles

Challenge:

Assume 8 visual encodings and n data fields

Pick the best encoding from the exponential number of possibilities $(n+1)^8$

Principle of Consistency:

The properties of the image (visual variables) should match the properties of the data

Principle of Importance Ordering:

Encode the most important information in the most effective way

Mackinlay's expressiveness criteria

Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express *all* the facts in the set of data, and *only* the facts in the data.

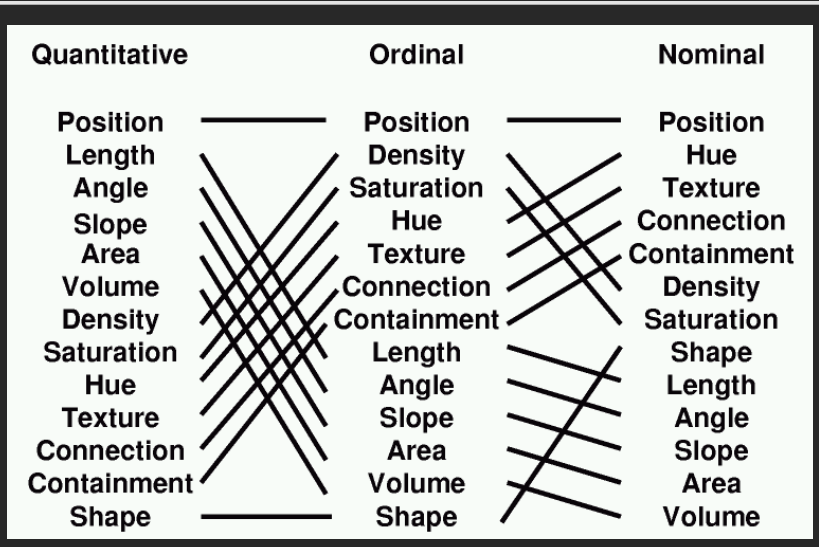
Mackinlay's effectiveness criteria

Effectiveness

A visualization is more effective than another visualization if the information conveyed by one visualization is more readily *perceived* than the information in the other visualization.

Subject of perception lecture

Mackinlay's ranking



Conjectured *effectiveness* of the encoding

Graphical Perception

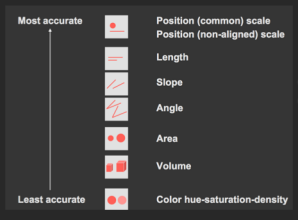
APT: Automatic Chart Construction

Automating the design of graphical presentation of relational information
J. Mackinlay, 1986

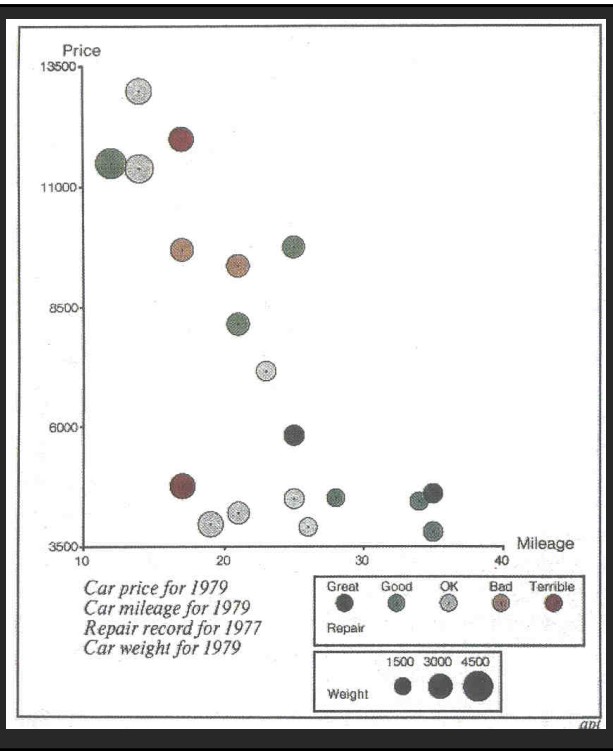
Encode most important data using highest ranking visual variable for the data type

Year	Exports	Imports
1700	170,000	300,000
1701	171,000	302,000
1702	176,000	303,000
...	...	...

→ **1. Year (Q)**
2. Exports (Q)
3. Imports (Q)



→ **mark: lines**
Year → x-pos (Q)
Exports → y-pos (Q)
Imports → y-pos (Q)



[Mackinlay, APT, 1986]

Announcements

Announcements

Class participation requirements

- Complete readings before class
- In-class discussion
- Post at least 1 discussion substantive comment/question by noon the day after lecture (short paragraph)

Office hours on website

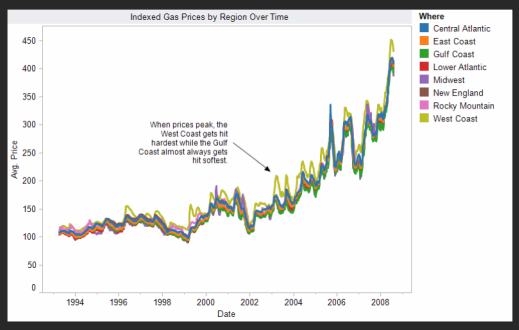
Class wiki
https://magrawala.github.io/cs448b-fa18

A2: Exploratory Data Analysis

Use Tableau to formulate & answer questions

First steps

- Step 1: Pick a domain
- Step 2: Pose questions
- Step 3: Find data
- Iterate



Create visualizations

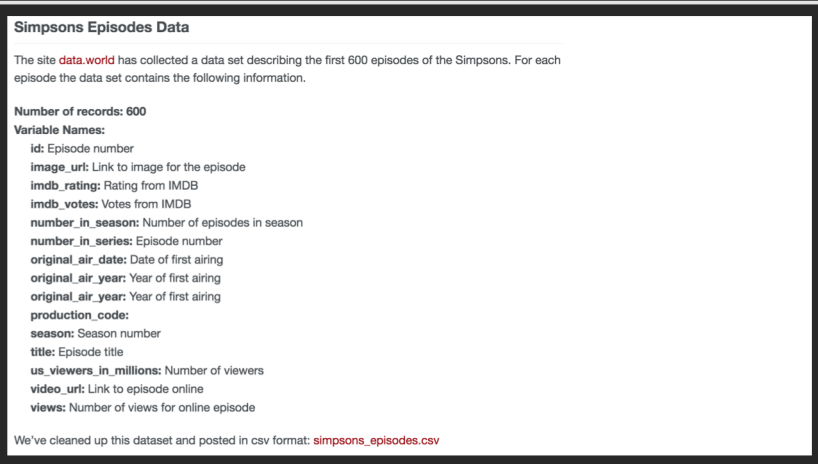
- Interact with data
- Question will evolve
- Tableau

Make wiki notebook

- Keep record of all steps you took to answer the questions

Due before class on Oct 15, 2018

A1: Visualization Design



Simpsons Episodes Data

The site data.world has collected a data set describing the first 600 episodes of the Simpsons. For each episode the data set contains the following information.

Number of records: 600
Variable Names:
- **id:** Episode number
- **image_url:** Link to image for the episode
- **imdb_rating:** Rating from IMDB
- **imdb_votes:** Votes from IMDB
- **number_in_season:** Number of episodes in season
- **number_in_series:** Episode number
- **original_air_date:** Date of first airing
- **original_air_year:** Year of first airing
- **original_air_year:** Year of first airing
- **production_code:**
- **season:** Season number
- **title:** Episode title
- **us_viewers_in_millions:** Number of viewers
- **video_url:** Link to episode online
- **views:** Number of views for online episode

We've cleaned up this dataset and posted in csv format: simpsons_episodes.csv

Simpsons Episodes

Due by noon today

Design Considerations

Expressiveness
- Do the mappings show the facts and only the facts?
- Are visual mappings consistent? (e.g., respect color mappings)

Effectiveness
- Are perceptually effective encodings used?
- Are the most important data mapped to the most effective visual variables?

Cognitive Load (Efficiency)
- Are there extraneous visual elements?

Data Transformation
- Are transformations (filter, sort, derive, aggregate) appropriate?

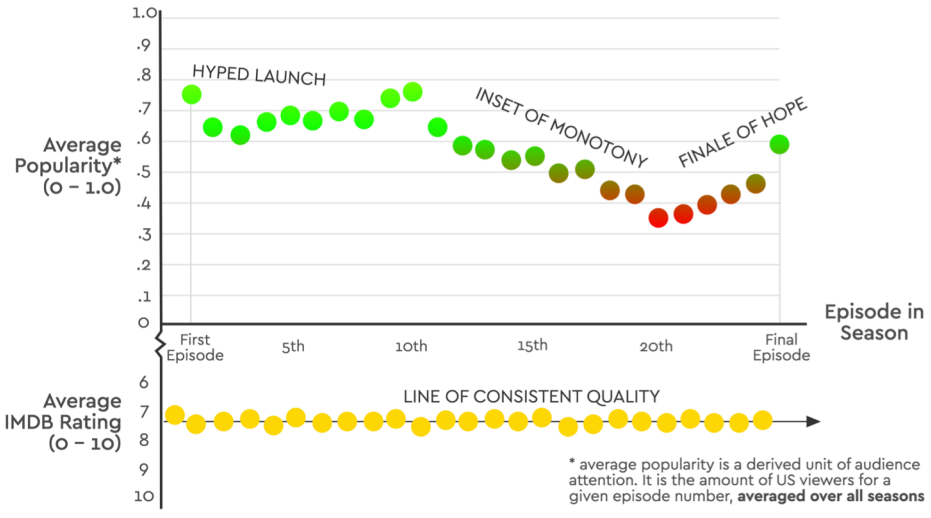
Guides (Non-Data Elements)
- Descriptive, consistent: Title, Label, Caption, Source, Annotations
- Meaningful references: Gridlines, Legend

Design Space of A1 Submissions

Spatial Encoding	**Scatterplot, Line Charts, some Bar charts**
Color Encoding	**Redundant with spatial dim** **Quantitative (dual encoding)**
Data Transformation	**Aggregations (box + whiskers)**
Labeling	**Title, Caption, Axis labels, Some Legends** **Some annotations**

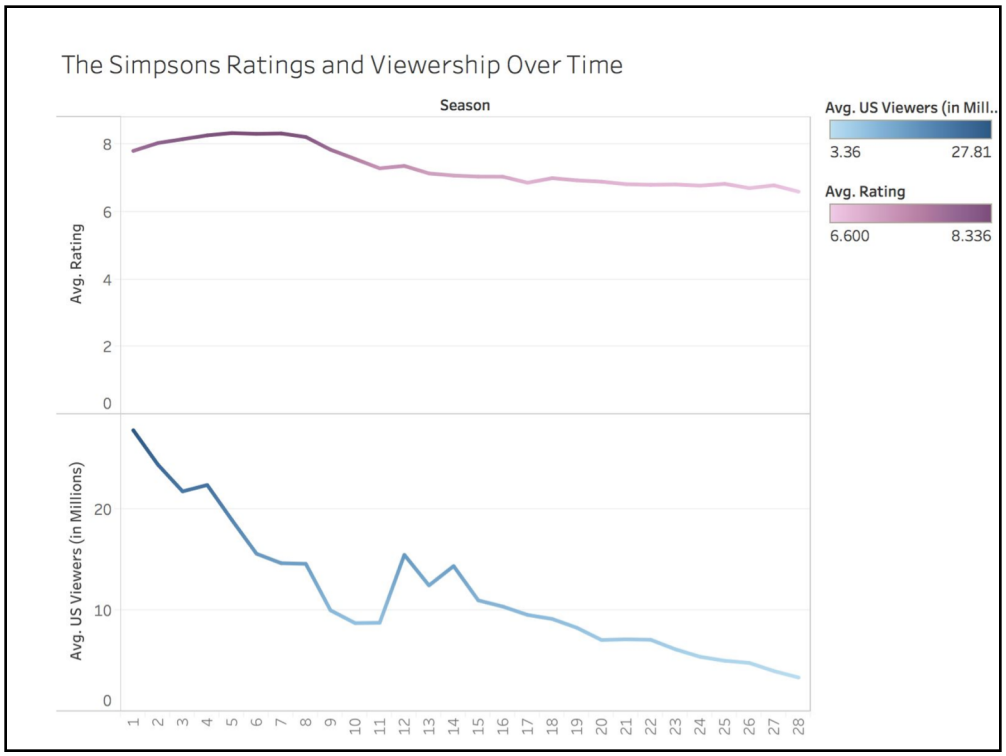
Viewership Movement Over a Season

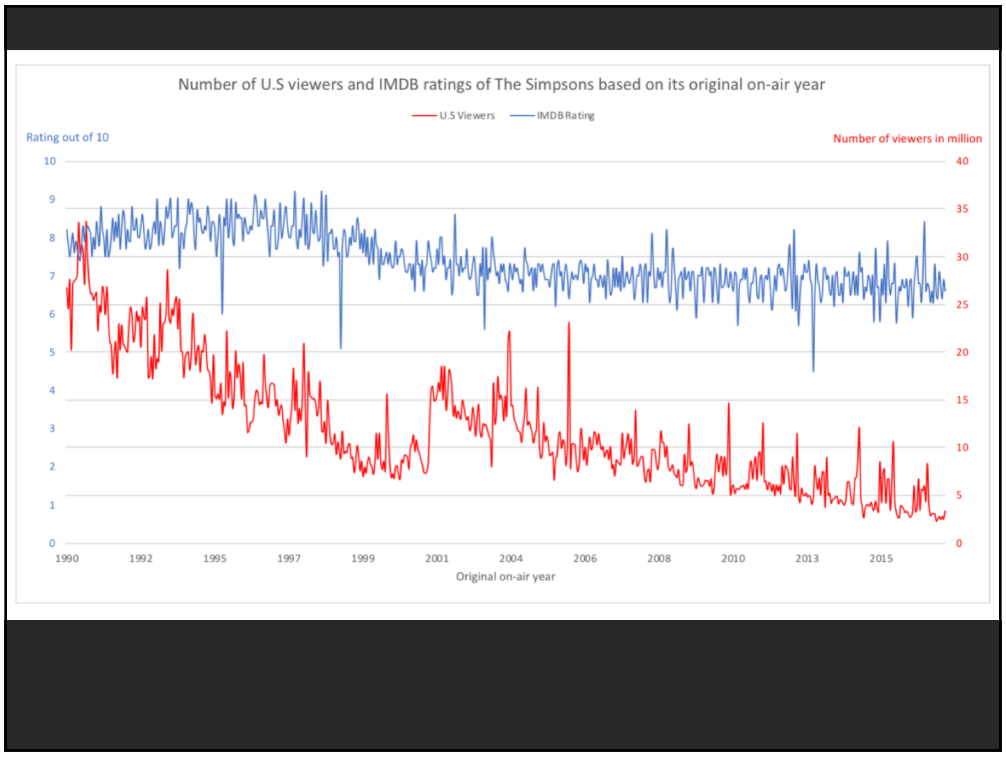
by Anand Upender

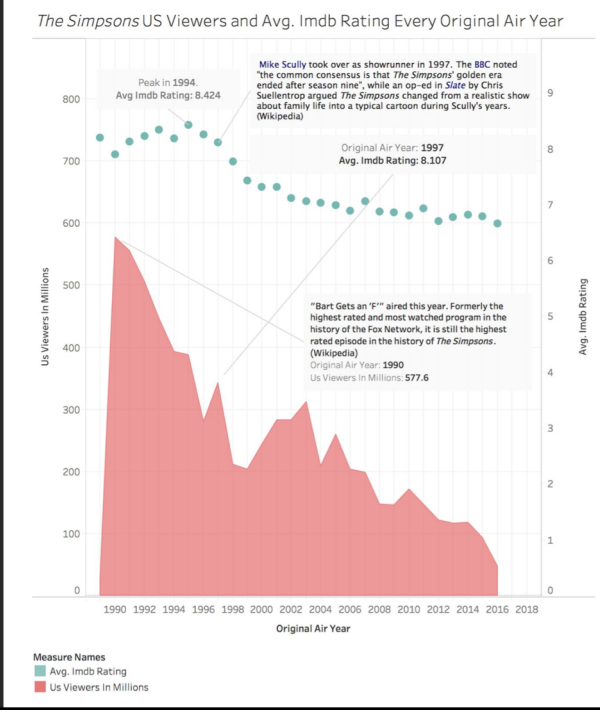


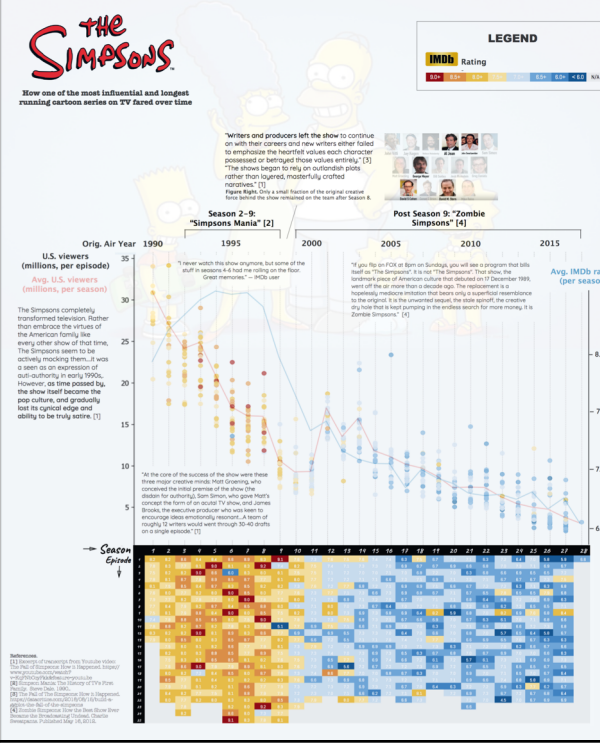
Graphed in Matplotlib. Touched up in Sketch.

Thanks to data.world


The Simpsons Ratings and Viewership Over Time
Season
Avg. US Viewers (in Mill..
3.36
27.81
Avg. Rating
6.600
8.336
Avg. Rating
Avg. US Viewers (in Millions)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28


Number of U.S viewers and IMDB ratings of The Simpsons based on its original on-air year
U.S Viewers
IMDB Rating
Rating out of 10
Number of viewers in million
1990 1992 1995 1997 1999 2001 2004 2006 2008 2010 2013 2015
Original on-air year


The Simpsons US Viewers and Avg. Imdb Rating Every Original Air Year
Peak in 1994.
Avg Imdb Rating: 8.424
Mike Scully took over as showrunner in 1997. The BBC noted "the common consensus is that The Simpsons' golden era ended after season nine", while an op-ed in Slate by Chris Suellentrop argued The Simpsons changed from a realistic show about family life into a typical cartoon during Scully's years. (Wikipedia)
Original Air Year: 1997
Avg. Imdb Rating: 8.107
"Bart Gets an 'F'" aired this year. Formerly the highest rated and most watched program in the history of the Fox Network, it is still the highest rated episode in the history of The Simpsons. (Wikipedia)
Original Air Year: 1990
Us Viewers In Millions: 577.6
Us Viewers in Millions
Avg. Imdb Rating
Original Air Year
Measure Names
Avg. Imdb Rating
Us Viewers In Millions


THE SIMPSONS
How one of the most influential and longest running cartoon series on TV fared over time
LEGEND
IMDb Rating
"Writers and producers left the show to continue on with their careers and new writers either failed to emphasize the heartfelt values each character possessed or betrayed those values entirely." [3]
"The shows began to rely on outlandish plots rather than layered, masterfully crafted narratives." [3]
Season 2–9: "Simpsons Mania" [2]
Post Season 9: "Zombie Simpsons" [4]
Orig. Air Year
U.S. viewers (millions, per episode)
Avg. U.S. viewers (millions, per season)
The Simpsons completely transformed television. Rather than embrace the virtues of the American family like every other show of that time, The Simpsons seem to be actively mocking them...It was a seen as an expression of anti-authority in early 1990s. However, as time passed by, the show itself became the pop culture, and gradually lost its cynical edge and ability to be truly satire. [1]
Avg. IMDb rating (per season)
"I never watch this show anymore, but some of the stuff in seasons 4-6 had me rolling on the floor. Great memories." — IMDb user
"If you flip on FOX at 8pm on Sundays, you will see a program that bills itself as "The Simpsons". It is not "The Simpsons". That show, the landmark piece of American culture that debuted on 17 December 1989, went off the air more than a decade ago. The replacement is a hopelessly mediocre imitation that bears only a superficial resemblance to the original. It is the unwanted sequel, the stale spinoff, the creative dry hole that is kept pumping in the endless search for more money. It is Zombie Simpsons." [4]
"At the core of the success of the show were these three major creative minds: Matt Groening, who conceived the initial premise of the show (the disdain for authority), Sam Simon, who gave Matt's concept the form of an acutal TV show, and James Brooks, the executive producer who was keen to encourage ideas emotionally resonant...A team of roughly 12 writers would went through 30-40 drafts on a single episode." [1]
Season
Episode

In-Class Review

Procedure

Break into groups of 3

Present your visualization – in order by last name – 3 min each to describe what your visualization shows, and design choices.

Others should write down critique on sheet

We will keep time and tell you to switch

Critique in order by last name – rubric on next slide (~5 min each)

- Tell author your critique.
- Give critiques to author

Author take photos of critiques and add to A1 along with a short response (1 paragraph) to the feedback.

In-Class Review

Expressiveness

- Do the mappings show the facts and only the facts?
- Are visual mappings consistent? (e.g., respect color mappings)

Effectiveness

- Are perceptually effective encodings used?
- Are the most important data mapped to the most effective visual variables?

Cognitive Load (Efficiency)

- Are there extraneous visual elements?

Data Transformation

- Are transformations (filter, sort, derive, aggregate) appropriate?

Guides (Non-Data Elements)

- Descriptive, consistent: Title, Label, Caption, Source, Annotations
- Meaningful references: Gridlines, Legend