INTRODUCTION TO VISUALIZATION (II)

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC

OUTLINE

- *Visualization. The basics*
- *History*
- **General Rules**
- Data, Tasks, Users
- Visualization as a Design Process
- The Visualization Mantra

GENERAL RULES

- Be honest
  - To yourself, to the audience
  - Did you check the data? Is it correct? Is it updated? Do you know how it was collected? The reason? Is there missing data?
- Above all, show the data
  - Is it necessary to add non informative elements? Does the data appear prominently? Are your visual encodings perceivable? By everybody? Can they be understood?

GENERAL RULES

- Explain encodings
  - Did you explain how are you using colors or shapes?
  - Are you using standard encodings? Did you create something new?
- Label your axes
  - Ensure they communicate what data or scales you are using
- Make the geometry change as the data
  - Data distortion leads to misinterpretations or disinformation

GENERAL RULES

- Take into account your audience
  - Physical conditions (e.g., color blindness), background (data literacy?)…
  - Is this to be shown in a big screen? A mobile?
  - How much the people knows about the data? About the problems? Are they domain experts or general public?
- Tell a story, give a message
  - Make your visualizations memorable
  - Communicating through stories is typically more engaging
  - Do you have a takeaway message? Write it

OUTLINE

- *Visualization. The basics*
- *History*
- General Rules
- **Data, Tasks, Users**
- Visualization as a Design Process
- The Visualization Mantra

DATA, TASKS, USERS

- **Data** Types
  - *Nominal* (unordered set of names)
    - Examples: Car manufacturers and countries
    - Only test for inequality are possible
  - *Ordinal* (ordered set of non-measureable data)
    - Examples: days of the week and rating scales
    - Tests for equality and direction (<,>,=, ≠)

DATA, TASKS, USERS

- **Data** Types
  - *Quantitative* (measured or simulated data)
    - Examples: physical measurements of height, weight, length
    - Full set of arithmetic operations possible

- Transformations
  - Nominal to ordinal data by introducing order, e.g., alphabetical
  - Quantitative to ordinal data by binning (*classification*)

DATA, TASKS, USERS

- **Data** (structure)

Structure	Examples
1-dimensional	Alphabetic lists, source code, texts/documents
2-dimensional	Planar or map data, photos
3-dimensional	Molecules, human body, buildings
Temporal	{start, finish}, e.g., medical records, project management, historical presentations
Multi-dimensional	N attributes -> points in n-dimensional space, e.g., relational databases
Tree	Hierarchies or tree structures, e.g., file directories, business organizations
Network	Connected as graphs, e.g., communications networks, social networks

DATA, TASKS, USERS

- Visualization **tasks** (e.g. from Shneiderman, 1996)
  - Overview
  - Zoom
  - Filter
  - Details-on-demand
  - Relate
  - History
  - Extract

DATA, TASKS, USERS

- **Tasks**. Overview:
  - Gain an overview of the entire collection.
  - **Strategies:** zoomed out views of each data type to see the entire collection plus an adjoining detail view

DATA, TASKS, USERS

- **Tasks**. Zoom:
  - Zoom in on items of interest
  - Users typically have an interest in some portion of a collection, and they need tools to enable them to control the zoom focus and the zoom factor
  - Smooth zooming helps users preserve their sense of position and context
  - Zooming could be on one dimension at a time by moving the zoom bar controls or by adjusting the size of the field-of -view box

DATA, TASKS, USERS

- **Tasks**. Filter:
  - Filter out uninteresting items. Dynamic queries applied to the items in the collection
    - One of the key ideas in information visualization
  - By allowing users to control the contents of the display, users can quickly focus on their interests by eliminating unwanted items
  - Sliders, buttons, or other control widgets coupled to rapid display update (less than 100 milliseconds) is the goal
    - Even when there are tens of thousands of displayed items

DATA, TASKS, USERS

- **Tasks**. Details-on-demand:
  - Select an item or group and get details when needed
  - Once a collection has been trimmed to a few dozen items it should be easy to browse the details about the group or individual items
  - The usual approach is to simply click on an item to get a pop-up window with values of each of the attributes

DATA, TASKS, USERS

- **Tasks**. Relate:
  - View relationships among items

DATA, TASKS, USERS

- **Tasks**. History:
  - Keep a history of actions to support undo, replay, and progressive refinement
  - It is rare that a single user action produces the desired outcome
  - Information exploration is inherently a process with many steps, so keeping the history of actions and allowing users to retrace their steps is important

DATA, TASKS, USERS

- **Tasks**. Extract:
  - Allow extraction of sub-collections and of the query parameters
  - Once users have obtained the item or set of items they desire, it would be useful to be able to extract that set and save it to a file in a format that would facilitate other uses such as sending by email, printing, graphing, or insertion into a statistical or presentation package

DATA, TASKS, USERS

- **Users**. Limitations:
  - Computational capacity (speed)
  - Human perceptual and cognitive capacity (memory)
  - Display capacity (vision)

DATA, TASKS, USERS

- **Users**. Computational capacity:
  - Computer time and memory are limited resources
  - If the visualization system needs to deliver interactive response, algorithms must be efficient enough
    - Better a fraction of a second than minutes or hours

DATA, TASKS, USERS

- **Users**. Human perceptual and cognitive capacity:
  - Memory and attention are finite resources

DATA, TASKS, USERS

- **Users**. Memory notoriously limited
  - Both for long-term recall and for shorter-term working memory
  - Surprisingly little information is internally stored in visual working memory
    - Vulnerable to *change blindness*
      - Large changes not noticed if attending to something else
- **Users**. Vigilance is a highly limited resource
  - Ability to perform visual search tasks degrades quickly
    - Far worse results after several hours than in the first few minutes

DATA, TASKS, USERS

- **Users**. Display capacity
  - Often not enough space for the desired result
    - Screen resolution is not large enough to show all desired information simultaneously.
  - The *information density* of a is a measure of the amount of information encoded versus the amount of unused space
    - Must carefully choose appropriate amount of information
    - Also referred to (at least partially) as Ink-ratio (see any of Tufte's books, e.g. "Envisioning information")

DATA, TASKS, USERS

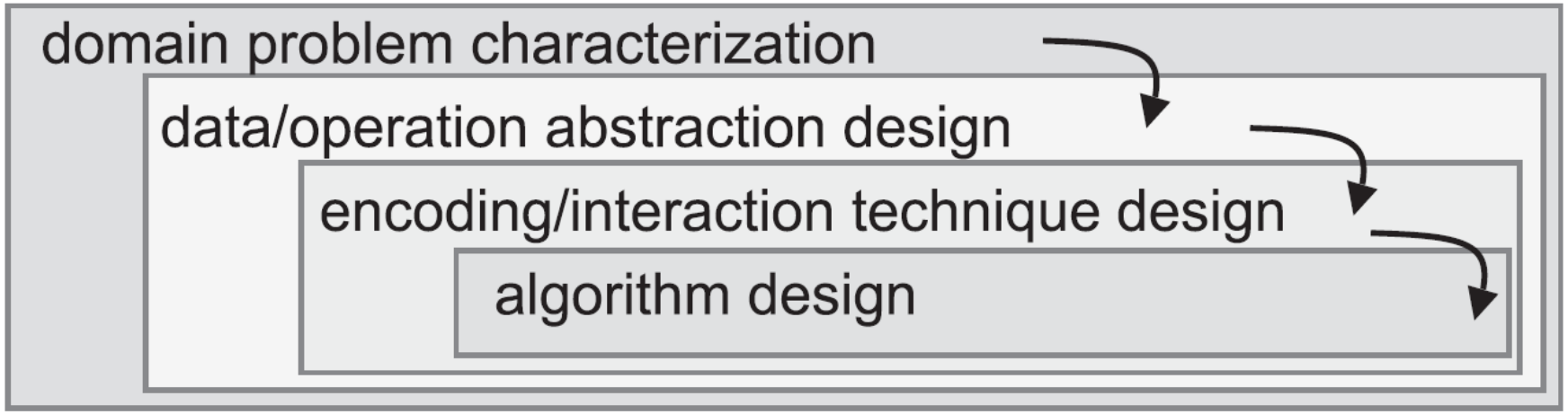
- **Users**. Display capacity. Tradeoffs information density:
  - Much information
    - Reduces the need for navigation
    - Reduces the need for exploration
    - User may be overwhelmed by visual clutter
  - White space
    - Tradeoff between clutter and wasting space, find sweet spot between dense and sparse

OUTLINE

- *Visualization. The basics*
- *History*
- General Rules
- *Data, Tasks, Users*
- **Visualization as a Design Process**
- The Visualization Mantra

VISUALIZATION AS A DESIGN PROCESS

- Visualization design can be split into a cascading set of layers



VISUALIZATION AS A DESIGN PROCESS

- Problem characterization:
  - A given dataset will surely have many possible visual encodings
  - Choosing the right visual encoding can be guided by the specific needs of some intended user
  - Different questions, or *tasks*, require very different visual encodings

VISUALIZATION AS A DESIGN PROCESS

- Problem characterization:
  - Understanding the requirements of some target audience is a tricky problem
    - One commonly works with a group of target users over time
    - Users know they need to somehow view their data
      - They cannot directly articulate their needs as clear-cut tasks
        - In terms of operations on data types
    - Iterative design including gathering information from the target is a must

VISUALIZATION AS A DESIGN PROCESS

- Data abstraction
  - Once the domain problem has been identified, it must be abstracted into a more generic representation
  - Problems of different domains may map to the same visualization abstraction
  - Operations can be performed for the abstraction
    - Generic operations: Sorting, filtering, characterizing trends and distributions, finding anomalies, correlation, outliers…
    - Specific operations: Following a path in a graph…
  - Abstraction involves transformations from the raw data to the derived dimensions
    - Often different type from the original data
      - Tabular data to a graph, graph to a tree…

VISUALIZATION AS A DESIGN PROCESS

- Technique and algorithm design:
  - Select appropriate visual encoding
  - Determine (& implement) adequate interaction techniques

VISUALIZATION AS A DESIGN PROCESS

- Validation:

  Each layer can be validated

  - Each of them have different validation requirements
  - **Problem domain:** Is the target audience performing tasks that would benefit from the visualization tool?
    - Different measures could be applied, but some of them (e. g. adoption rate) can only be evaluated once the tool has been developed
  - **Abstraction:** Are the developed data types usable to solve the problem.
    - A field study might assess whether the audience is using the tool and how.
    - Images might be analysed quantitatively and qualitatively.

VISUALIZATION AS A DESIGN PROCESS

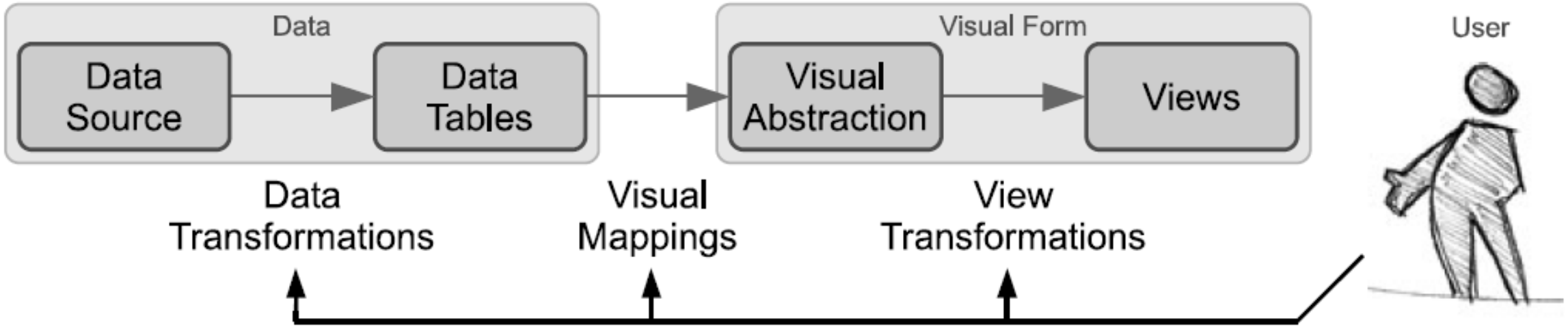
- Validation
  - **Visual encoding:** Do the visual encoding communicate effectively the abstraction?
    - Ensure no perceptual and cognitive principles are violated.
    - Formal studies can also be carried out, take measures
  - **Algorithm/technique design:** Verify the designed algorithm to visualize and render is faster, takes less memory
    - Could analyse & compare computational complexity
    - Time in performance can be measured

VISUALIZATION AS A DESIGN PROCESS

- Visualization pipeline
  - Answers to the question: How is it visualized?
  - Defines the steps to obtain the final visualization
    - Raw data need to be transformed into image data in a data-dependent and task-specific manner
    - Map data to geometry and corresponding visual attributes like color, position, size, or shape, also called **visual variables**

VISUALIZATION AS A DESIGN PROCESS

- The visualization pipeline is dynamic (Aigner et al., 2011)



OUTLINE

- *Visualization. The basics*
- *History*
- General Rules
- *Data, Tasks, Users*
- *Visualization as a Design Process*
- **The Visualization Mantra**

THE VISUALIZATION MANTRA

- Schneiderman’s Mantra

1. Overview first
2. Zoom and filter
3. Then Details on Demand

THE VISUALIZATION MANTRA

- Schneiderman's Mantra

1. Overview first: Provide **big picture** of the data, no details
2. Zoom and filter
3. Then Details on Demand

THE VISUALIZATION MANTRA

- Schneiderman's Mantra

1. *Overview first: Provide big picture of the data, no details*
2. Zoom and filter: Focus on a **particular area** of the data
3. Then Details on Demand

THE VISUALIZATION MANTRA

- Schneiderman's Mantra

1. *Overview first: Provide big picture of the data, no details*
2. *Zoom and filter: Focus on a particular area of the data*
3. Then Details on Demand: Only when requested, **details of single data items**

THE VISUALIZATION MANTRA

- Keim et al's Visual Analytics Mantra (Keim et al., 2008)

1. Analyze first
2. Show the important
3. Zoom, filter, and analyze further
4. Details on Demand

OTHER RESOURCES

- T. Munzner's presentation (with the visualization definition): https://www.youtube.com/watch?v=Bc7Ajlud0is
- Minard's Napoleon March analyzed by R. Kosara: https://www.youtube.com/watch?v=hlb1uM_SOcE
- Florence Nightingale's Famous Rose Chart, by R. Kosara: https://www.youtube.com/watch?v=JZh8tUy_bnM

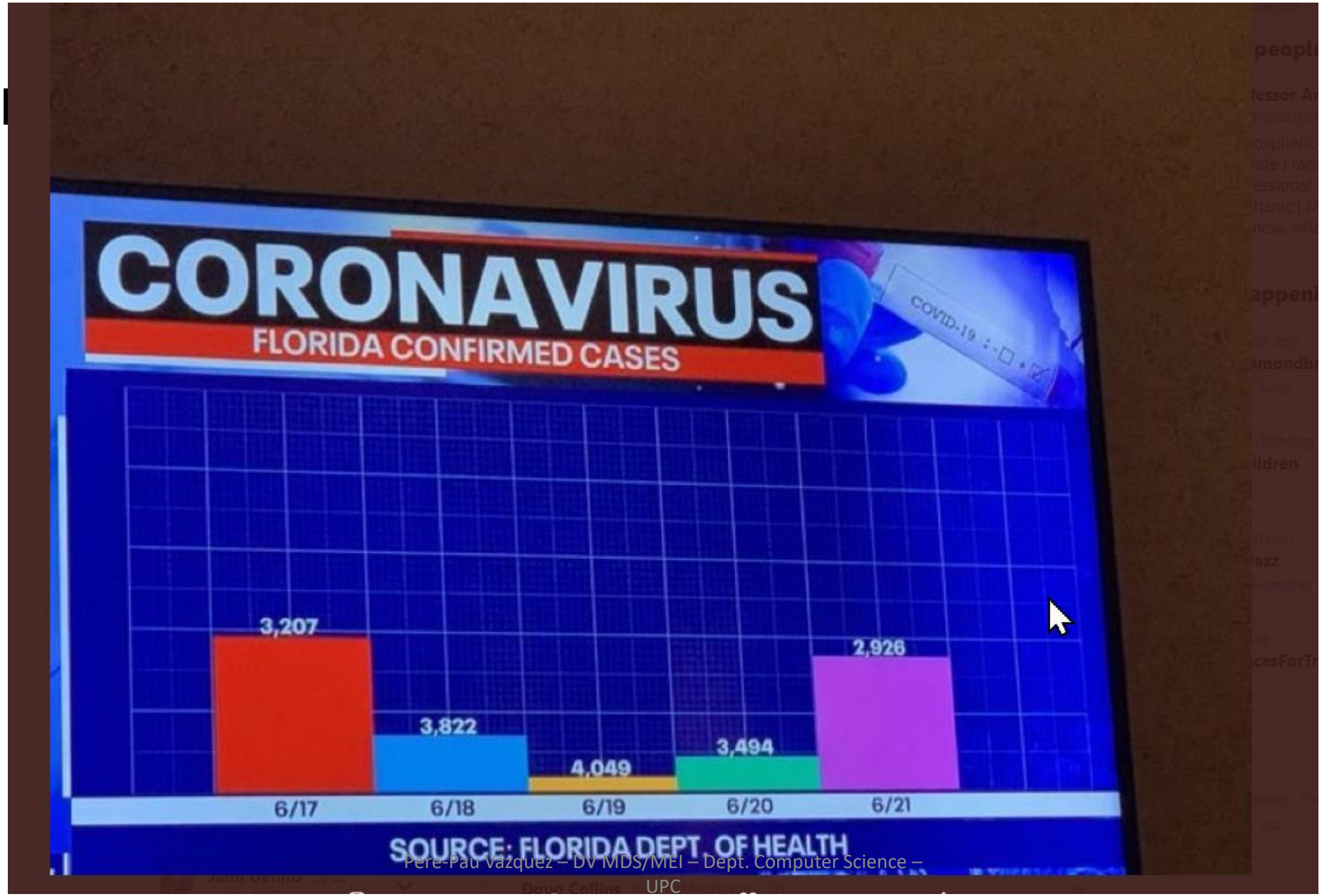
INTRODUCTION TO VISUALIZATION

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC

EXERCISE

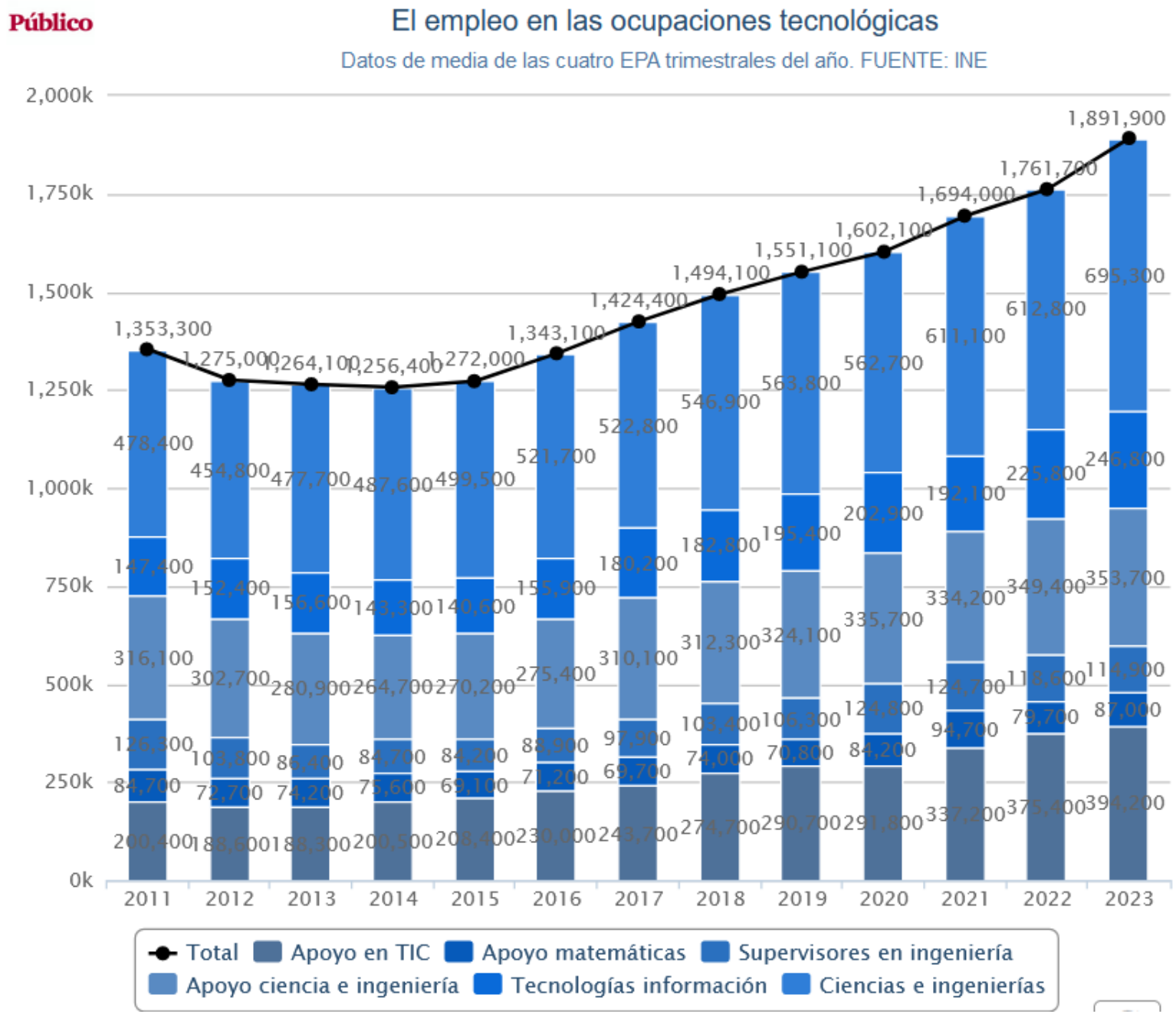
- The names of iPhones (e.g., iPhone 12, iPhone 12 Pro, iPhone 13…) can be seen as a variable…
  A. Quantitative, as their screen size can be measured in inches
  B. Qualitative, since they have no order
  C. Ordered, since there is an intrinsic order between them
  D. Ordered, because I can impose an order on them

EXE

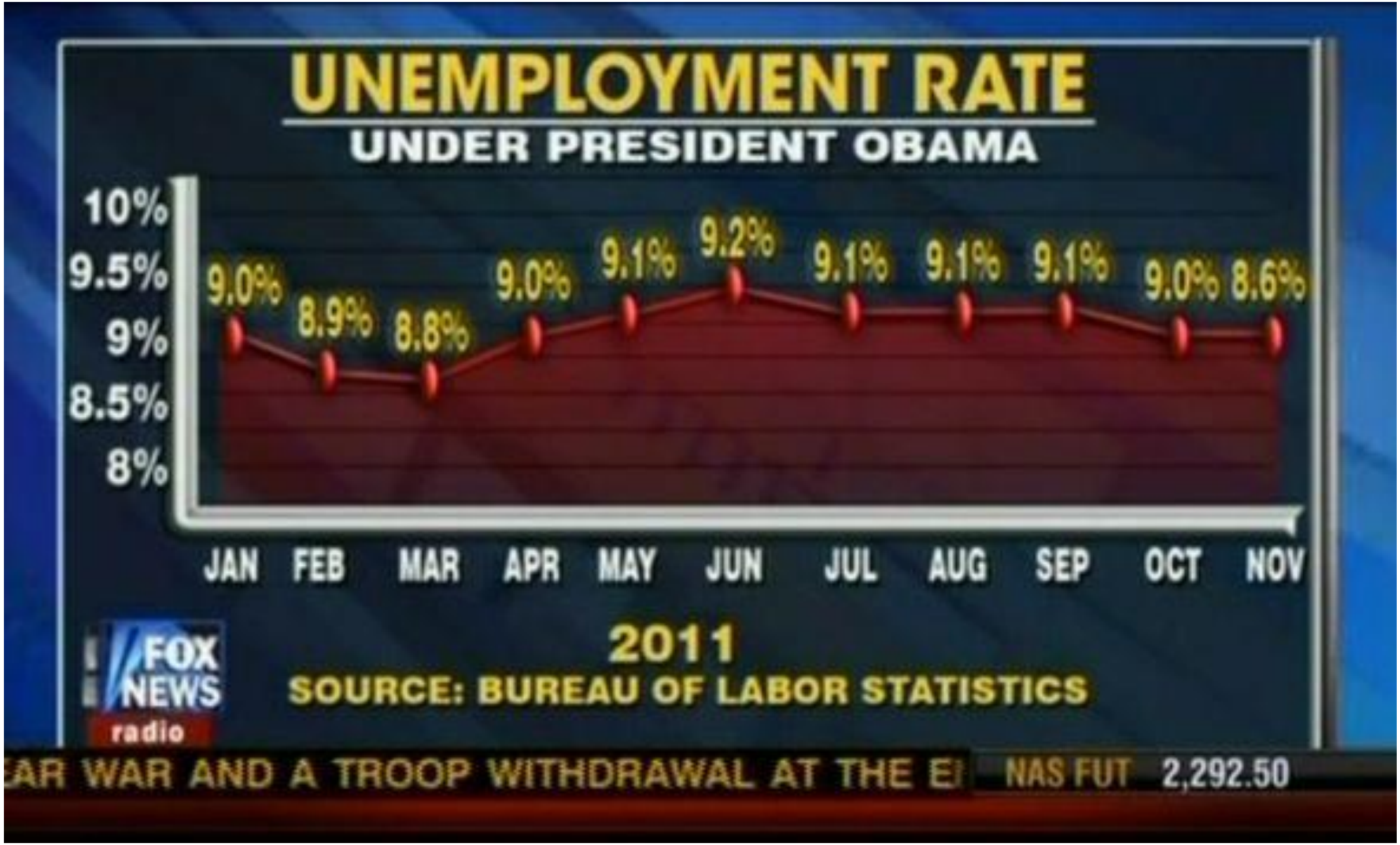

CORONAVIRUS
FLORIDA CONFIRMED CASES
COVID-19
3,207
3,822
4,049
3,494
2,926
6/17
6/18
6/19
6/20
6/21
SOURCE: FLORIDA DEPT. OF HEALTH

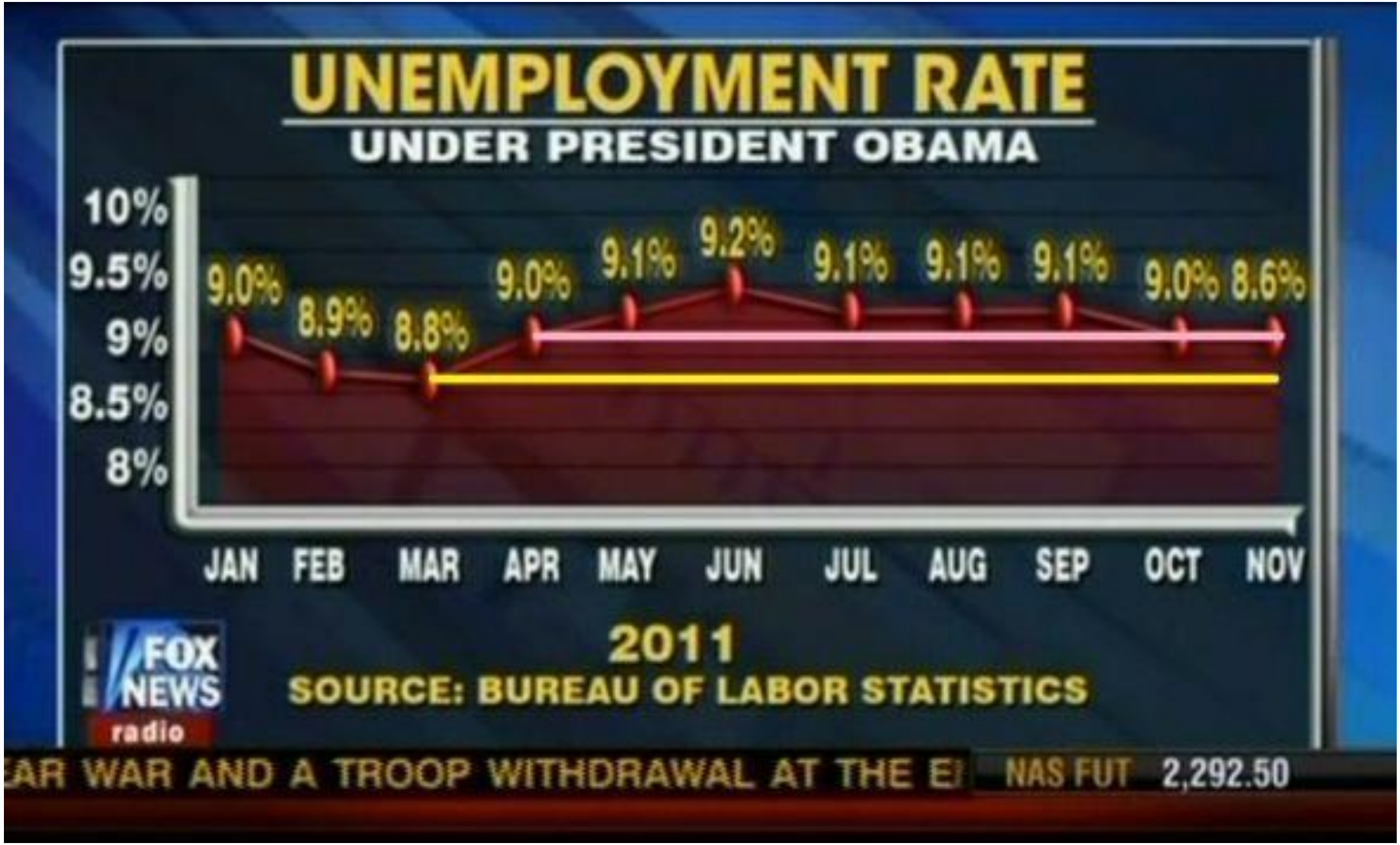
EXERCISE

- We have designed a color palette named pantone. Each color has a numerical code. What kind of variable can the color be?
  A. Quantitative, because numbers are quantitative
  B. Ordered, because there is a code that can be used to set <,> relationships between them
  C. Nominal, because colors have no intrinsic order

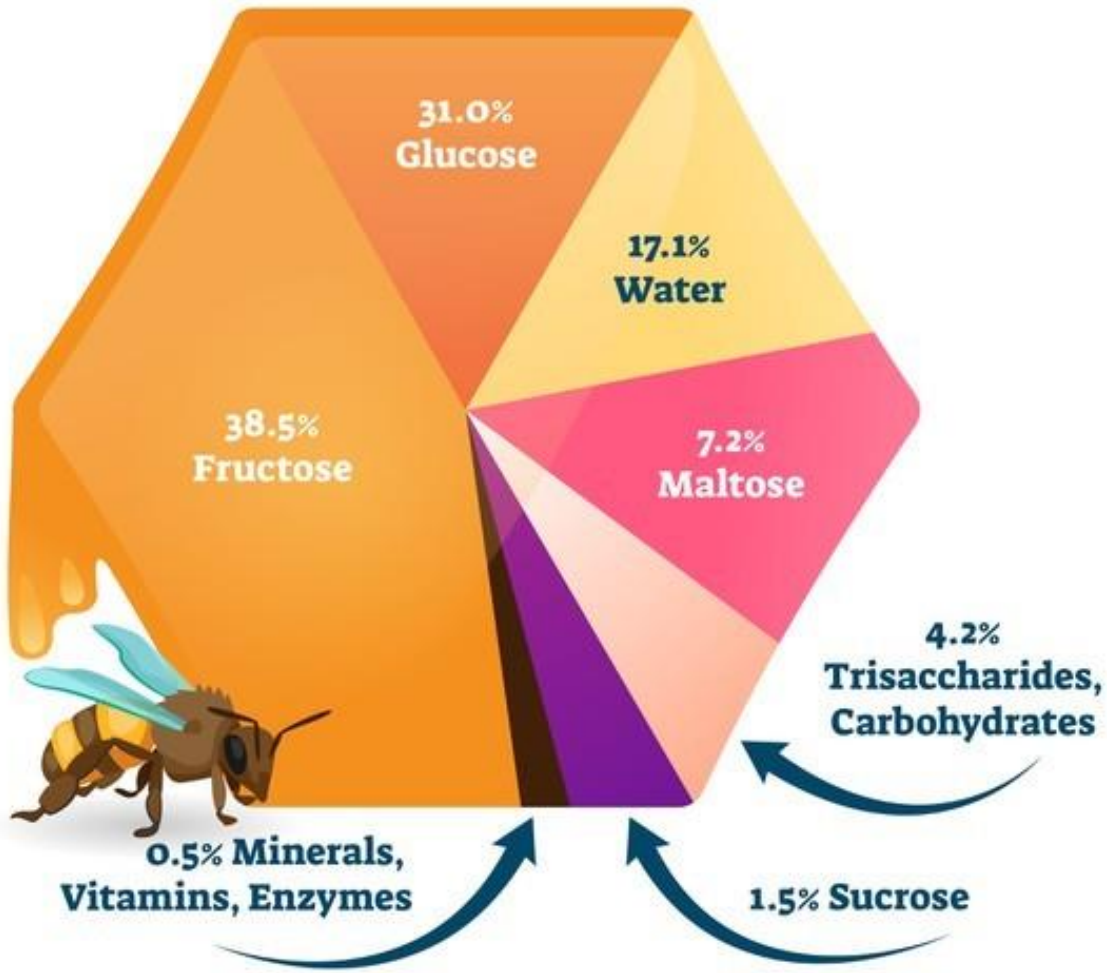


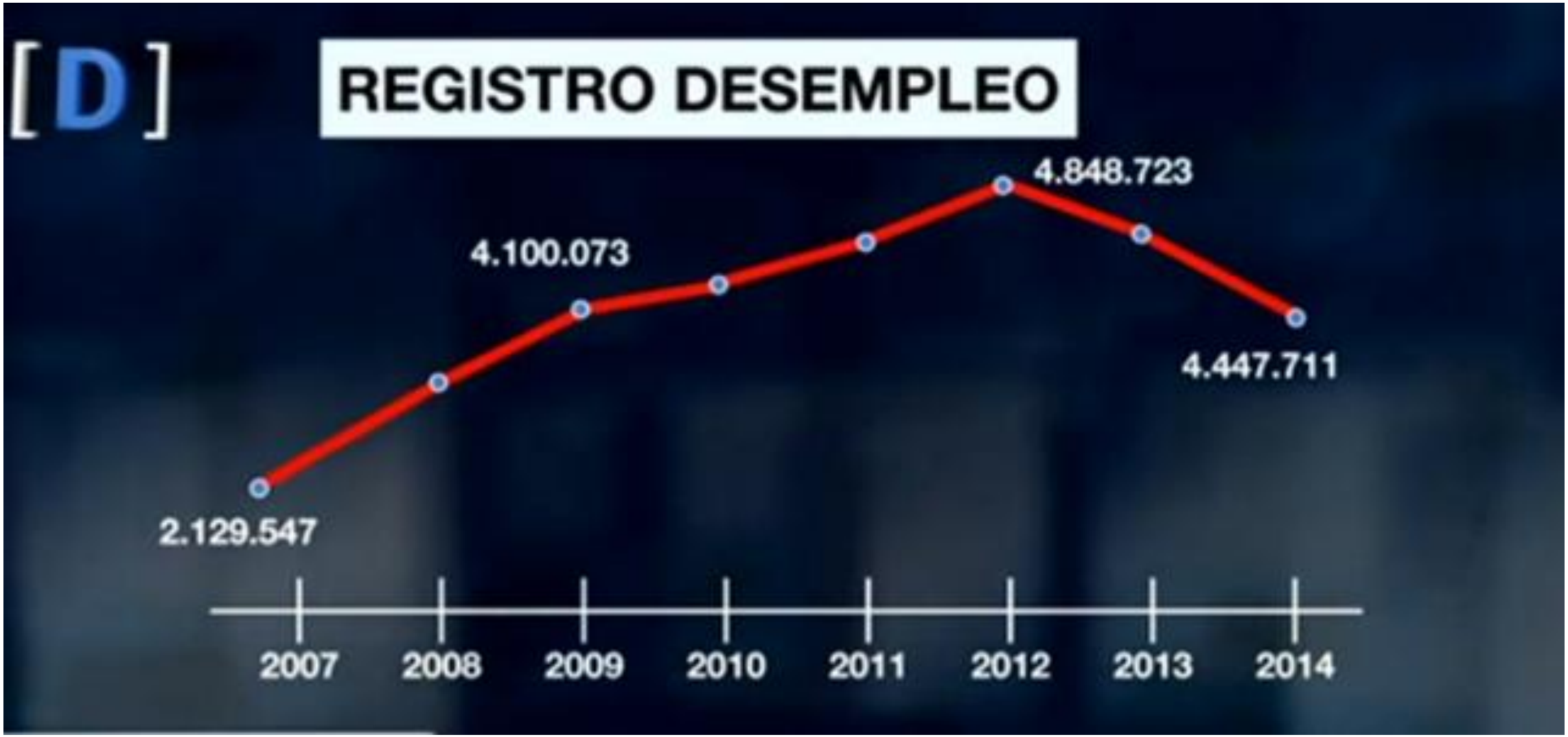
https://www.publico.es/economia/tecnologico-crece-precario-tres-categorias-no-alcanza-smi.html#md=modulo-portada-bloque:2col-t1;mm=mobile-medium

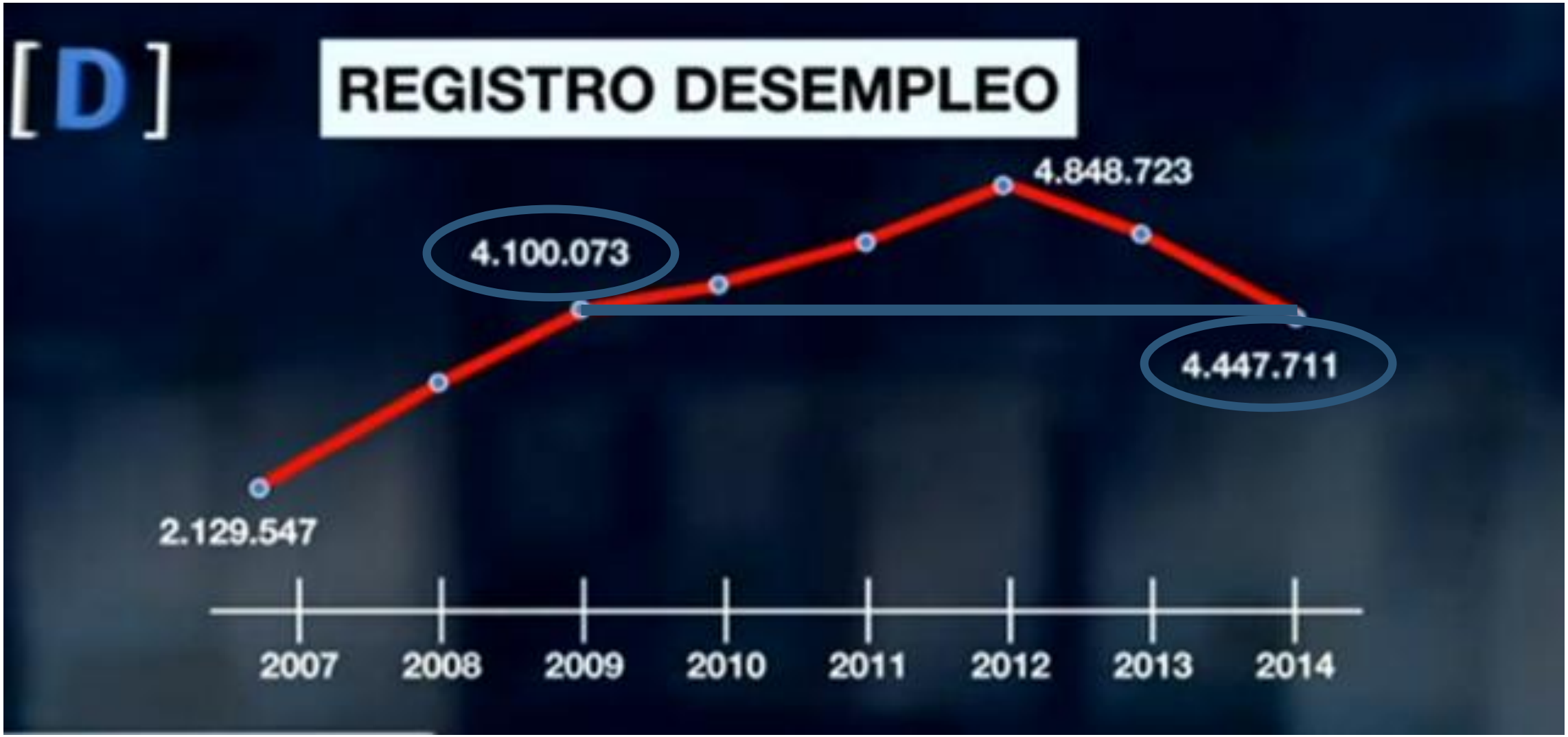

UNEMPLOYMENT RATE
UNDER PRESIDENT OBAMA
10%
9.5%
9%
8.5%
8%
9.0%
8.9%
8.8%
9.0%
9.1%
9.2%
9.1%
9.1%
9.1%
9.0%
8.6%
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
2011
SOURCE: BUREAU OF LABOR STATISTICS
FOX NEWS radio
AR WAR AND A TROOP WITHDRAWAL AT THE E
NAS FUT 2,292.50


UNEMPLOYMENT RATE
UNDER PRESIDENT OBAMA
10%
9.5%
9%
8.5%
8%
9.0%
8.9%
8.8%
9.0%
9.1%
9.2%
9.1%
9.1%
9.1%
9.0%
8.6%
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
2011
SOURCE: BUREAU OF LABOR STATISTICS
FOX NEWS radio
AR WAR AND A TROOP WITHDRAWAL AT THE E
NAS FUT 2,292.50

COMPOSITION OF HONEY

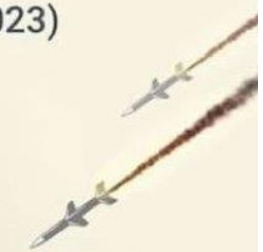



[D]
REGISTRO DESEMPLEO
4.848.723
4.100.073
4.447.711
2.129.547
2007
2008
2009
2010
2011
2012
2013
2014


[D]
REGISTRO DESEMPLEO
4.848.723
4.100.073
4.447.711
2.129.547
2007
2008
2009
2010
2011
2012
2013
2014

Israel-Palestinian Live Lost

Israel & Palestinian deaths documented by the UN (Jan 2008 - Sep 2023)



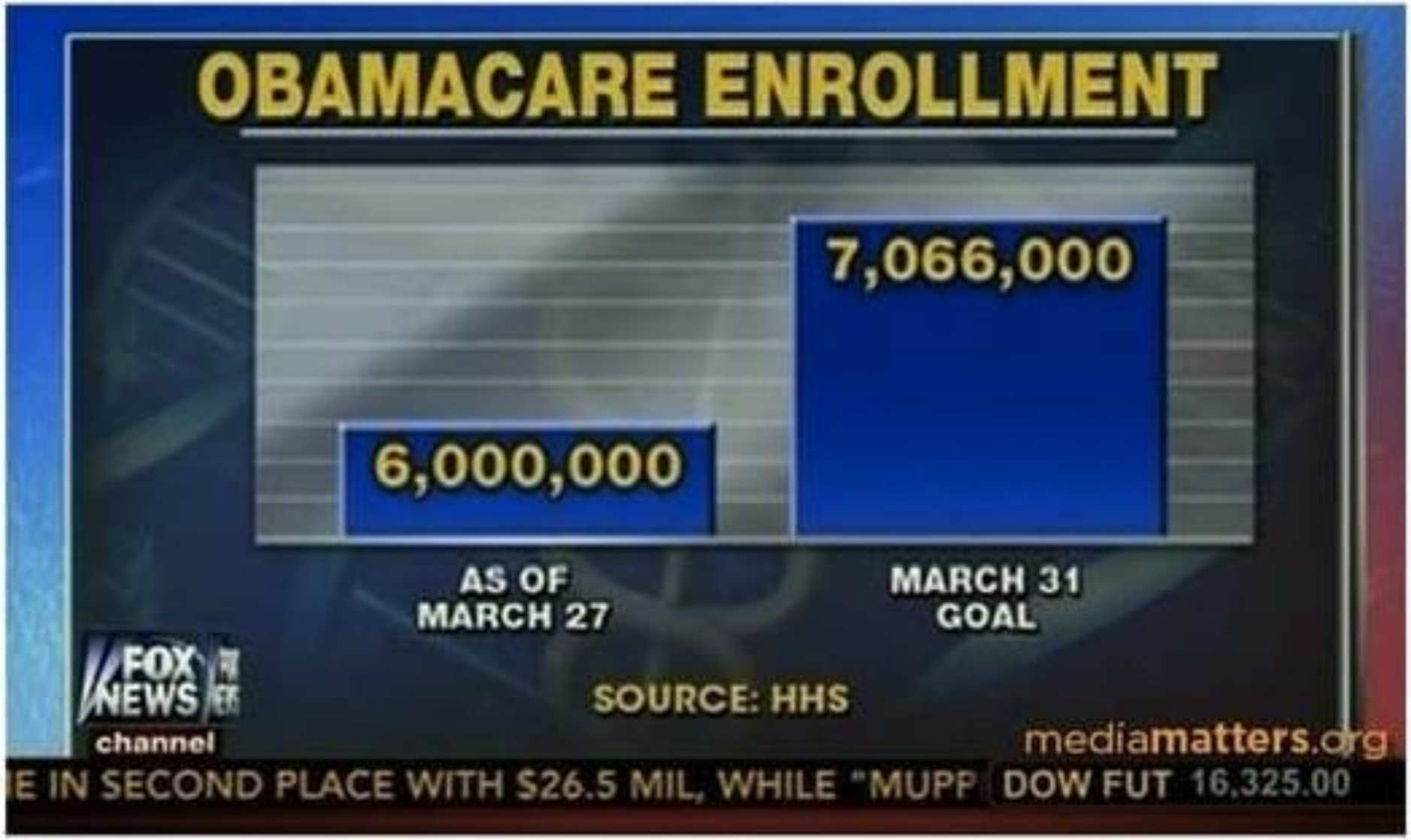
Palestinian Deaths	Year	Israel Deaths
899	2008	33
1,066	2009	11
95	2010	8
124	2011	11
260	2012	7
39	2013	6
2,329	2014	88
174	2015	26
109	2016	12
77	2017	17
300	2018	13
138	2019	12
30	2020	3
349	2021	11
191	2022	21
360	2023	30

Total Palestinian Deaths
6,540

Total Israeli Deaths
309

Data from 24 Jan 2008 to 30th October 2023
Source: United Nations Office for Coordination of Humanitarian Affairs

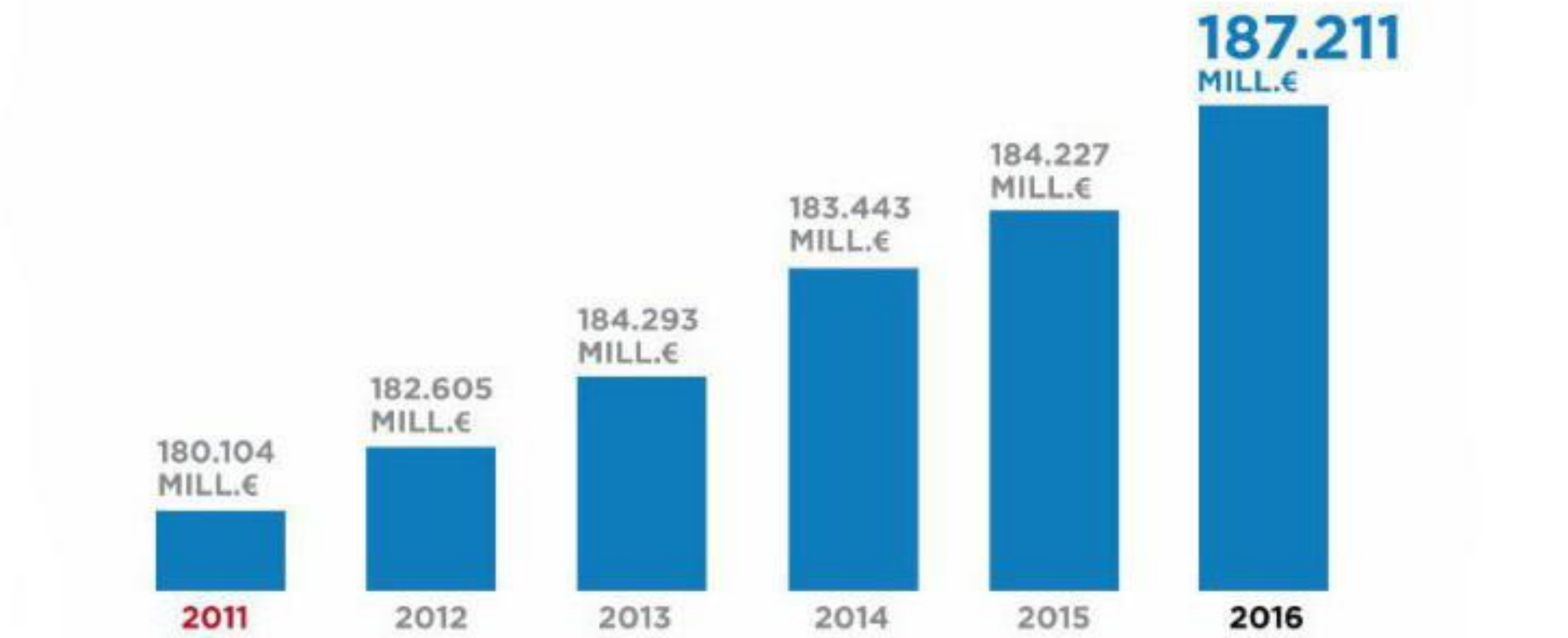
Obamacare Signups, According to Fox News



Source: Media Matters of America

GASTO SOCIAL

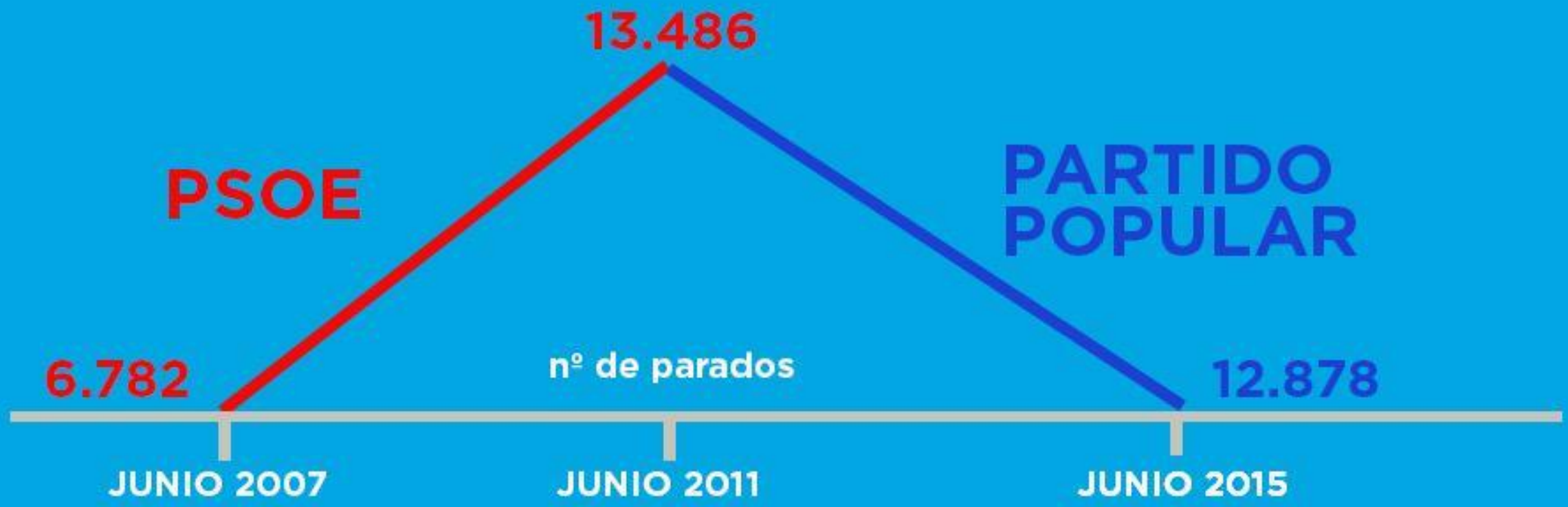
EDUCACIÓN, SANIDAD Y PROTECCIÓN SOCIAL



GASTO SOCIAL

EDUCACIÓN, SANIDAD Y PROTECCIÓN SOCIAL

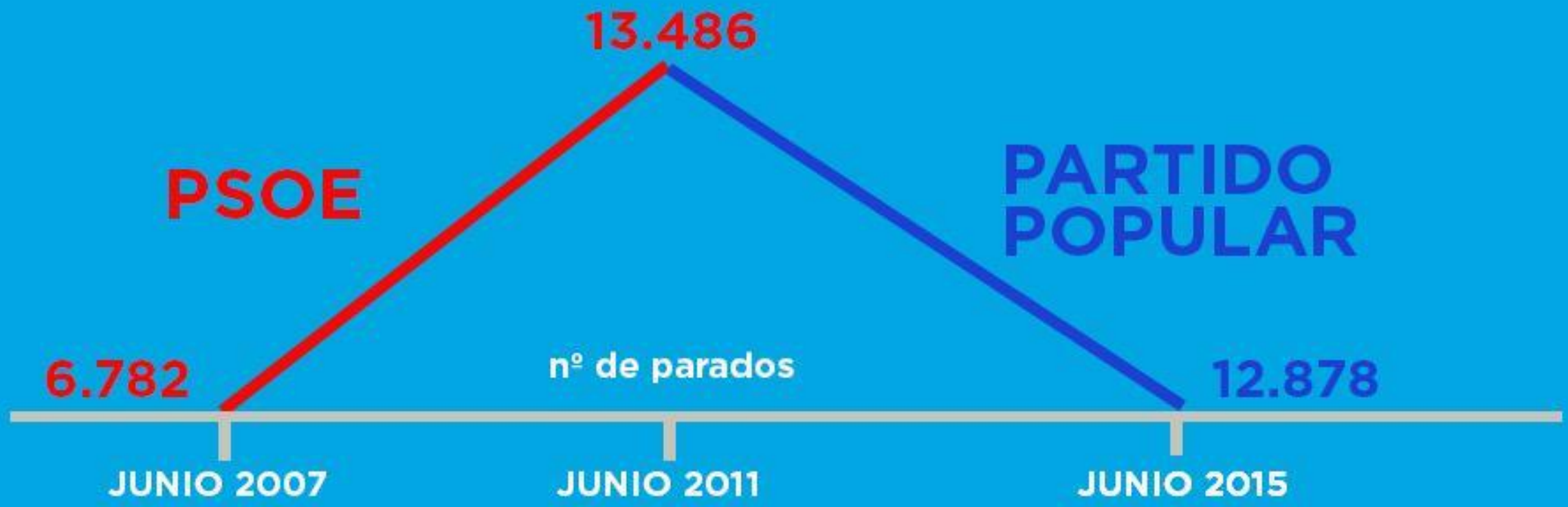
En Alcorcón ahora se crea empleo



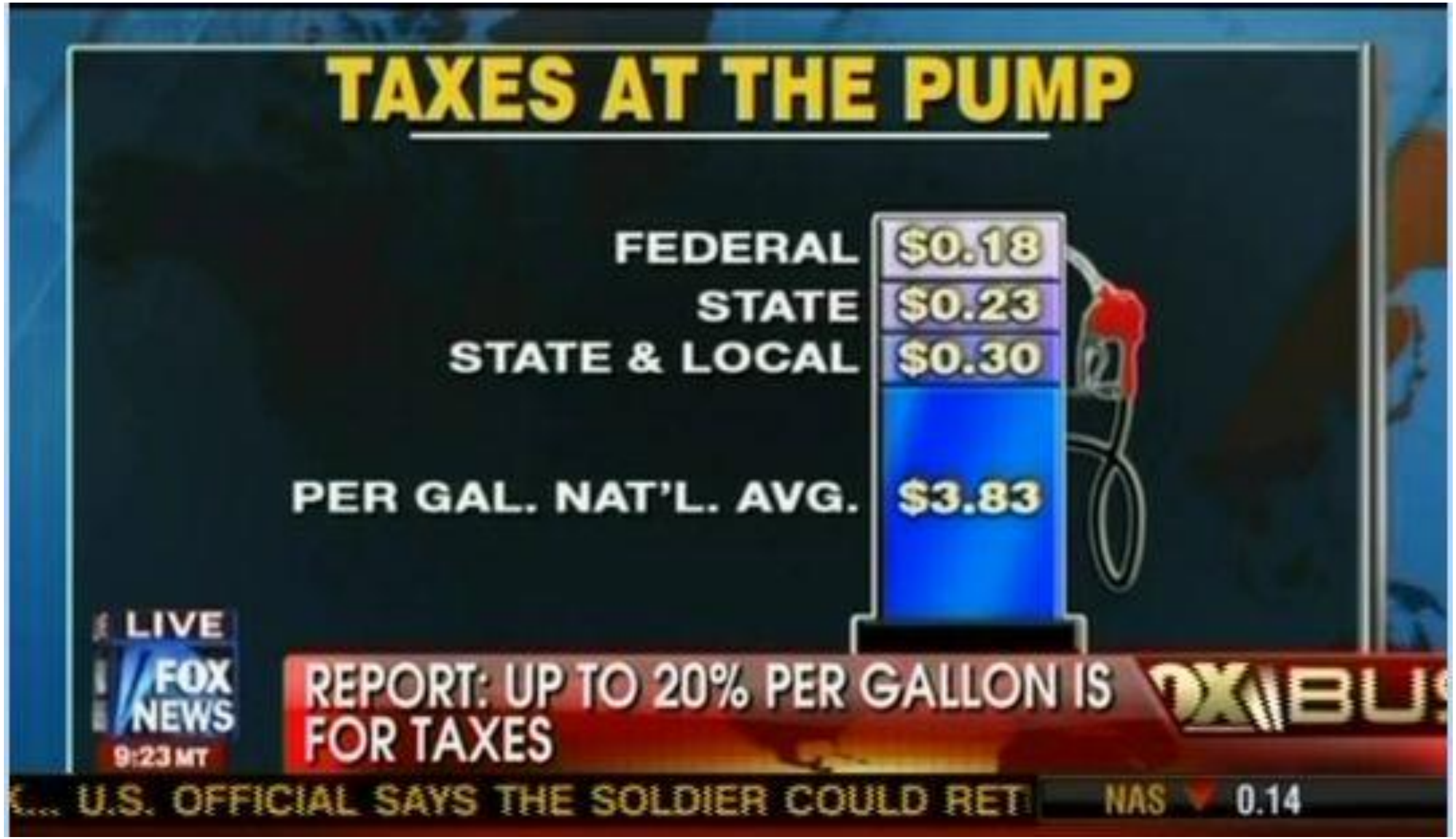
PP POPULARES de Alcorcón


SWEET MOTHER OF GOD

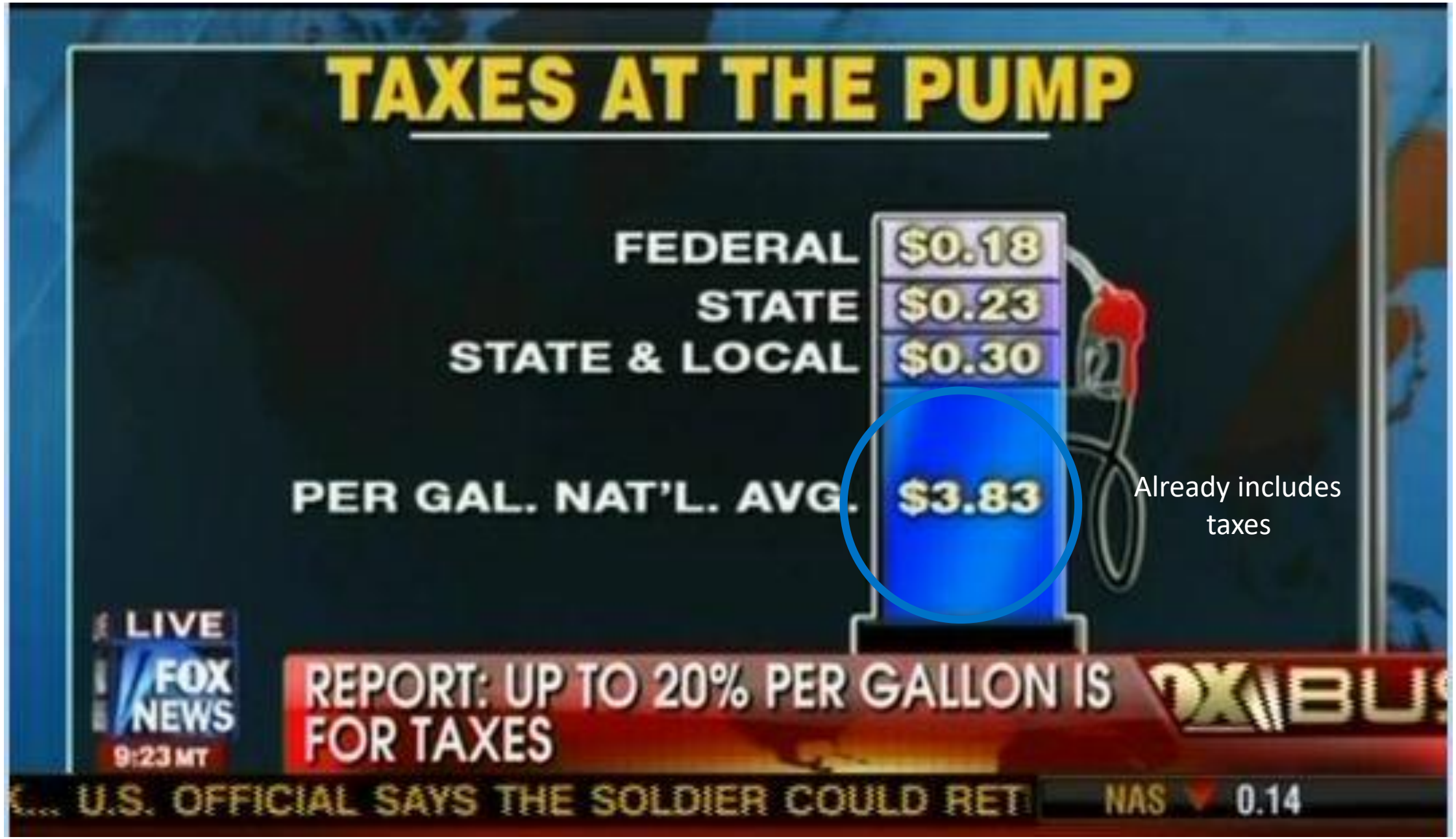
En Alcorcón ahora se crea empleo

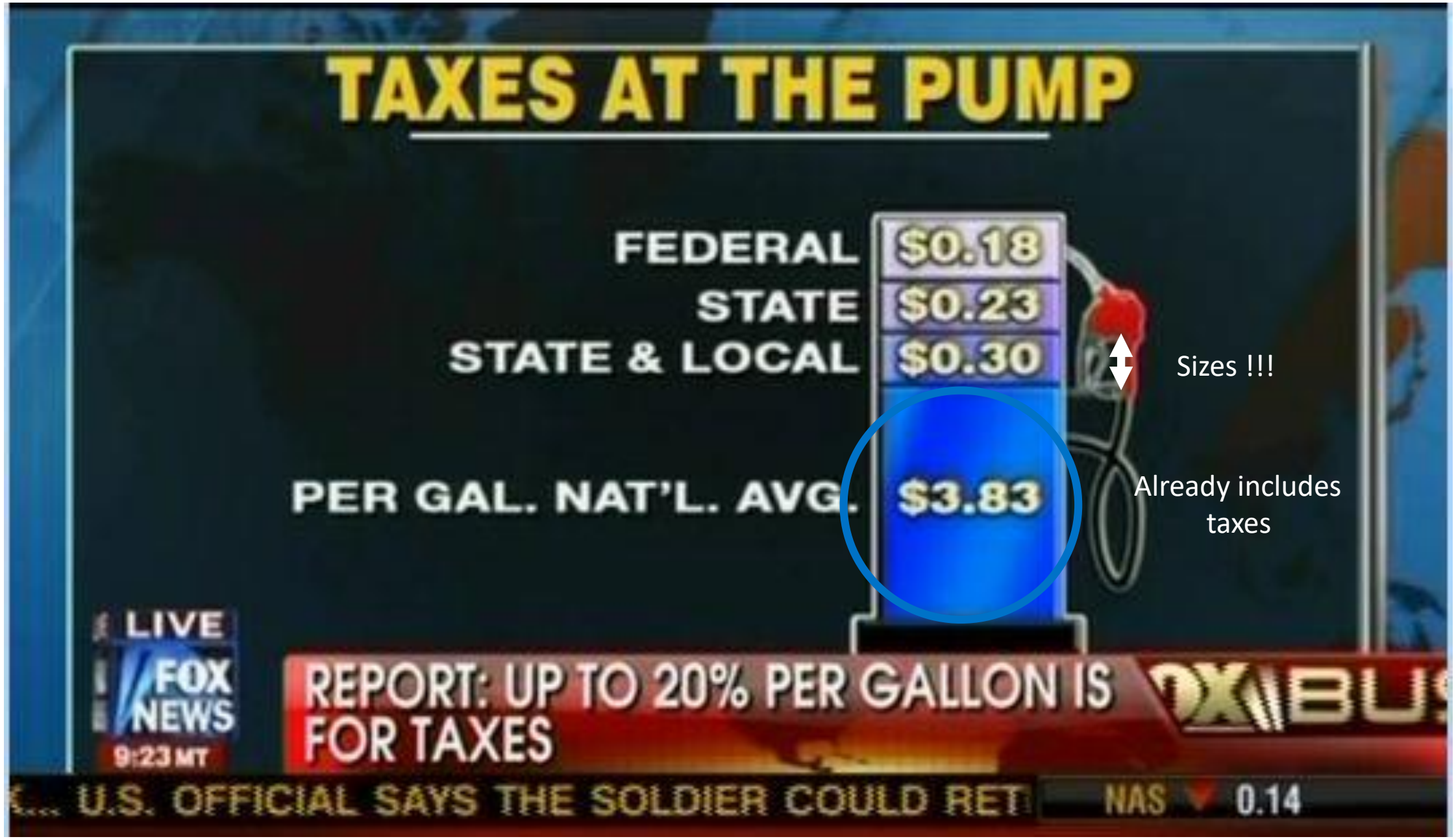


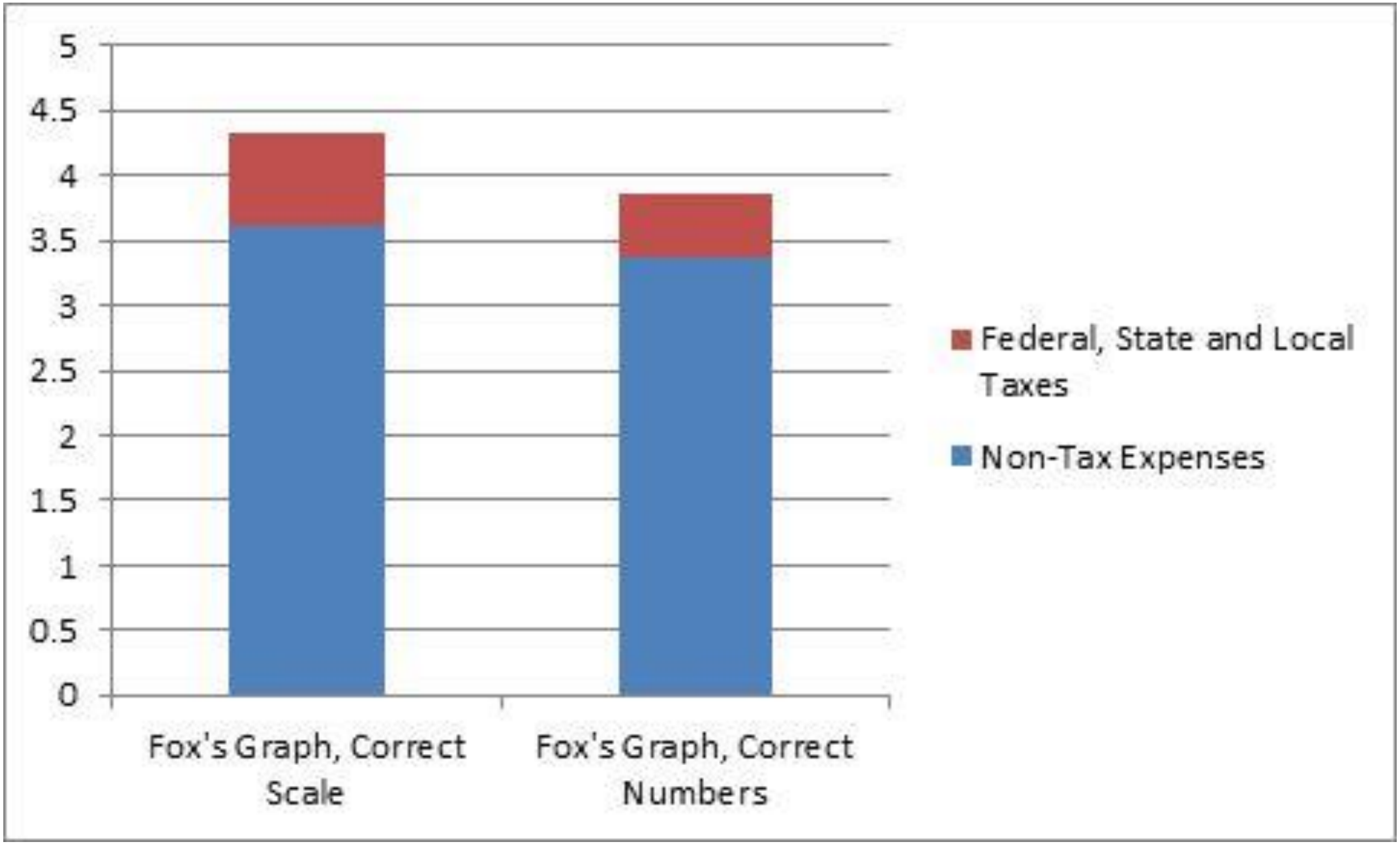
PP POPULARES de Alcorcón


TAXES AT THE PUMP
FEDERAL $0.18
STATE $0.23
STATE & LOCAL $0.30
PER GAL. NAT'L. AVG. $3.83
LIVE
FOX NEWS
9:23 MT
REPORT: UP TO 20% PER GALLON IS FOR TAXES
U.S. OFFICIAL SAYS THE SOLDIER COULD RET
NAS 0.14

WHERE IS THE TRICK? AKA “WHERE IS THE POOP”?


TAXES AT THE PUMP
FEDERAL $0.18
STATE $0.23
STATE & LOCAL $0.30
PER GAL. NAT'L. AVG. $3.83
Already includes taxes
LIVE
FOX NEWS
9:23 MT
REPORT: UP TO 20% PER GALLON IS FOR TAXES
U.S. OFFICIAL SAYS THE SOLDIER COULD RET
NAS 0.14

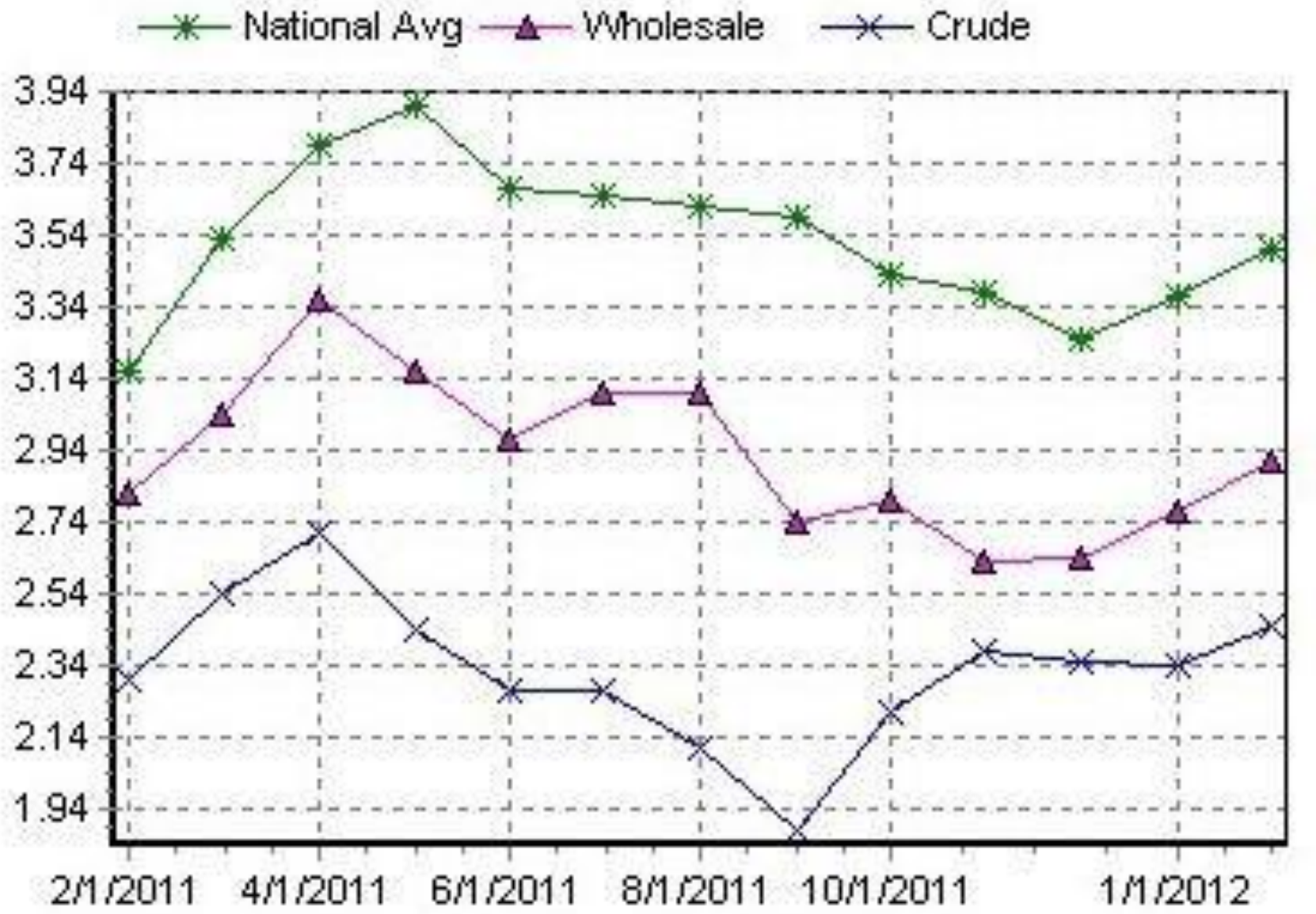

TAXES AT THE PUMP
FEDERAL $0.18
STATE $0.23
STATE & LOCAL $0.30
PER GAL. NAT'L. AVG. $3.83
Sizes !!!
Already includes taxes
LIVE
FOX NEWS
9:23 MT
REPORT: UP TO 20% PER GALLON IS FOR TAXES
U.S. OFFICIAL SAYS THE SOLDIER COULD RET
NAS 0.14


5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Fox's Graph, Correct Scale
Fox's Graph, Correct Numbers
Federal, State and Local Taxes
Non-Tax Expenses


COST OF GAS
NATIONAL AVERAGE
$3.60
$3.50
$3.40
$3.30
$3.20
$3.10
$3.00
$2.90
$3.17
$3.51
$3.57
LAST YEAR
LAST WEEK
CURRENT
FOX NEWS channel
SOURCE: AAA FUEL GAUGE REPORT
PAY... THE STRIKE, WHICH BEGAN LAST WEEK, IS
DOW ▲ 45.79

WHERE IS THE TRICK? AKA “WHERE IS THE POOP”?

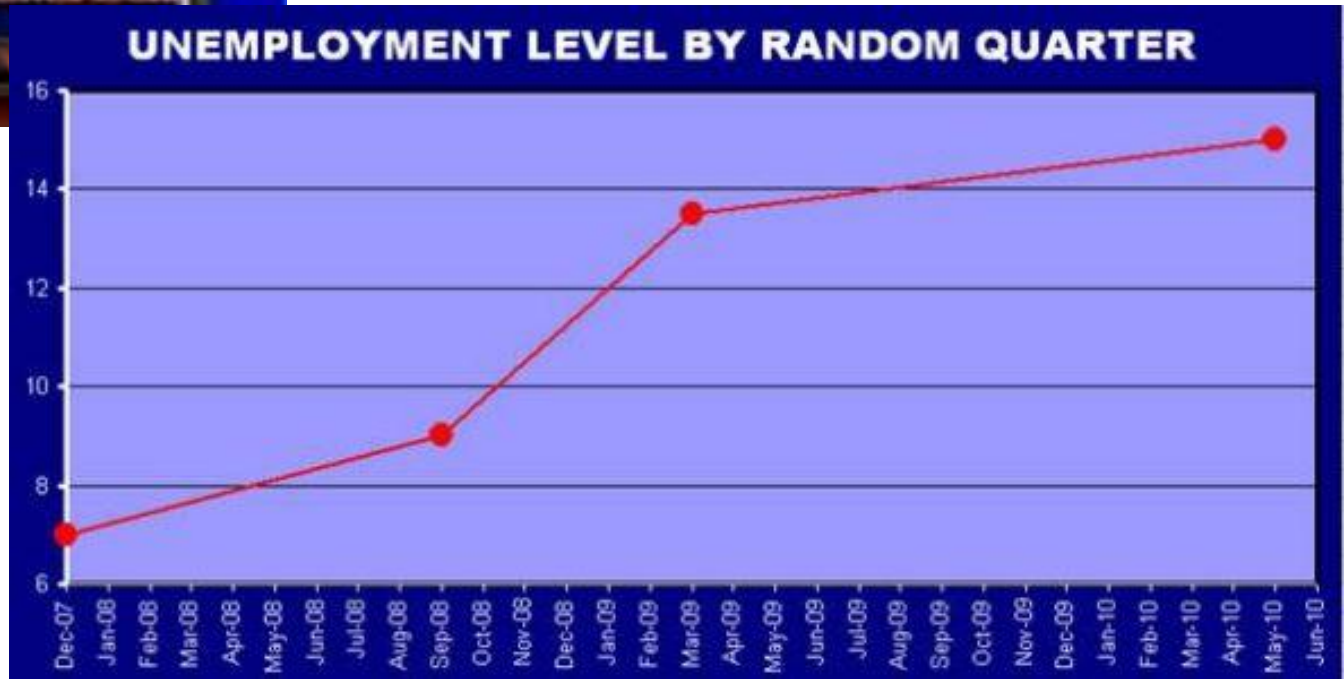
Timeline, data cherry picking (see next)

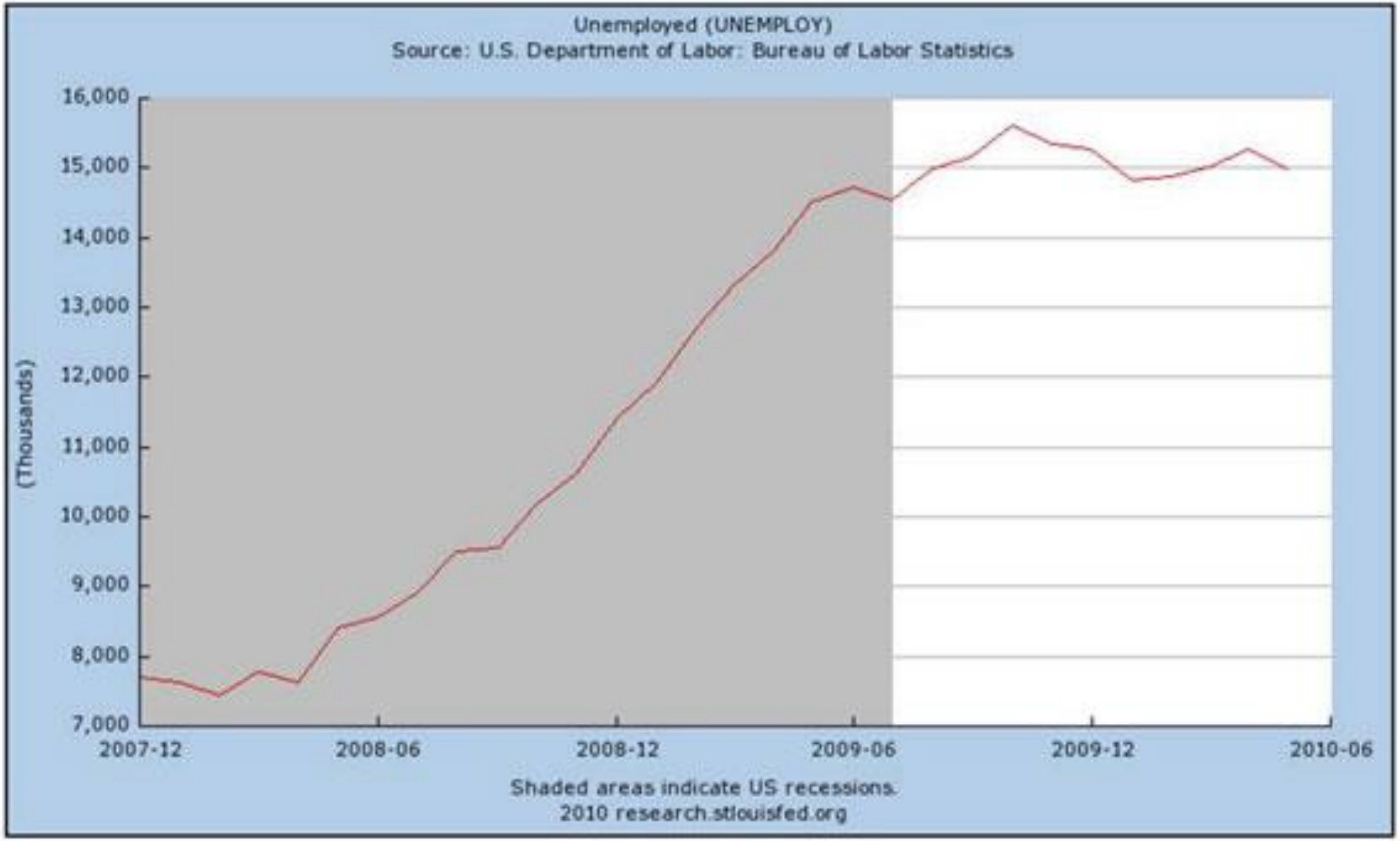


Real data over a year

Wrong scale!
ot job losses!

Same random months, correct scale




Unemployed (UNEMPLOY)
Source: U.S. Department of Labor: Bureau of Labor Statistics
16,000
15,000
14,000
13,000
12,000
11,000
10,000
9,000
8,000
7,000
(Thousands)
2007-12
2008-06
2008-12
2009-06
2009-12
2010-06
Shaded areas indicate US recessions.
2010 research.stlouisfed.org

EXERCISES

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC

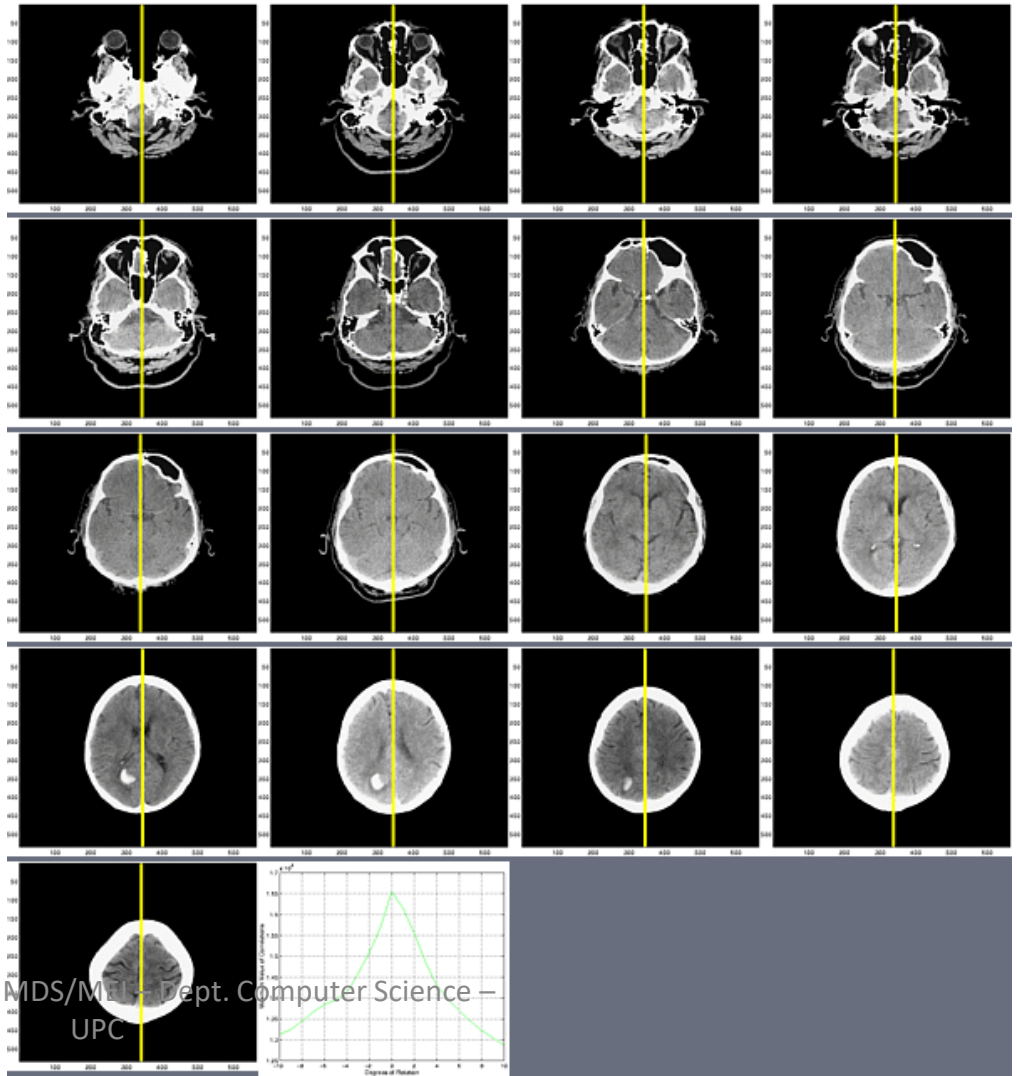
VISUALIZATION AREAS

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC

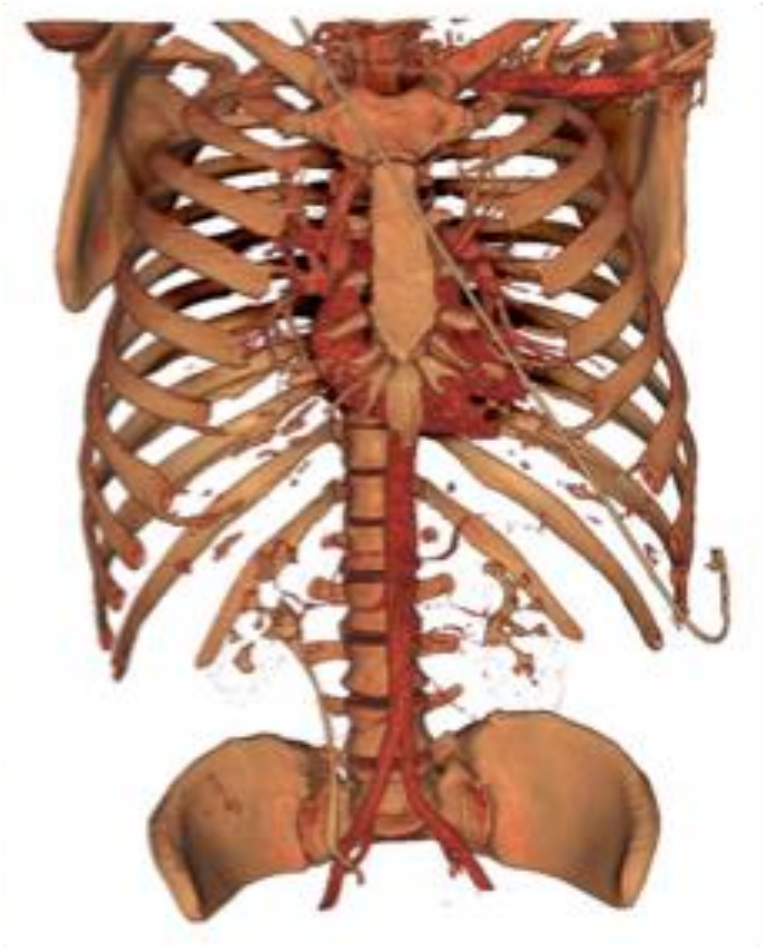
VISUALIZATION AREAS

- Based on data:
  - BioMed: Medical, biological, flows, astronomical data, molecular physics, seismic data, epidemiological data, neurons…
  - Geographical: maps, networks, goods/connection flows
  - Abstract: Graphs, genome analysis, health data
  - Economy: business analytics
  - Sports: sports analytics
  - Visual perception analysis
  - Explainable AI…
  - … and many others

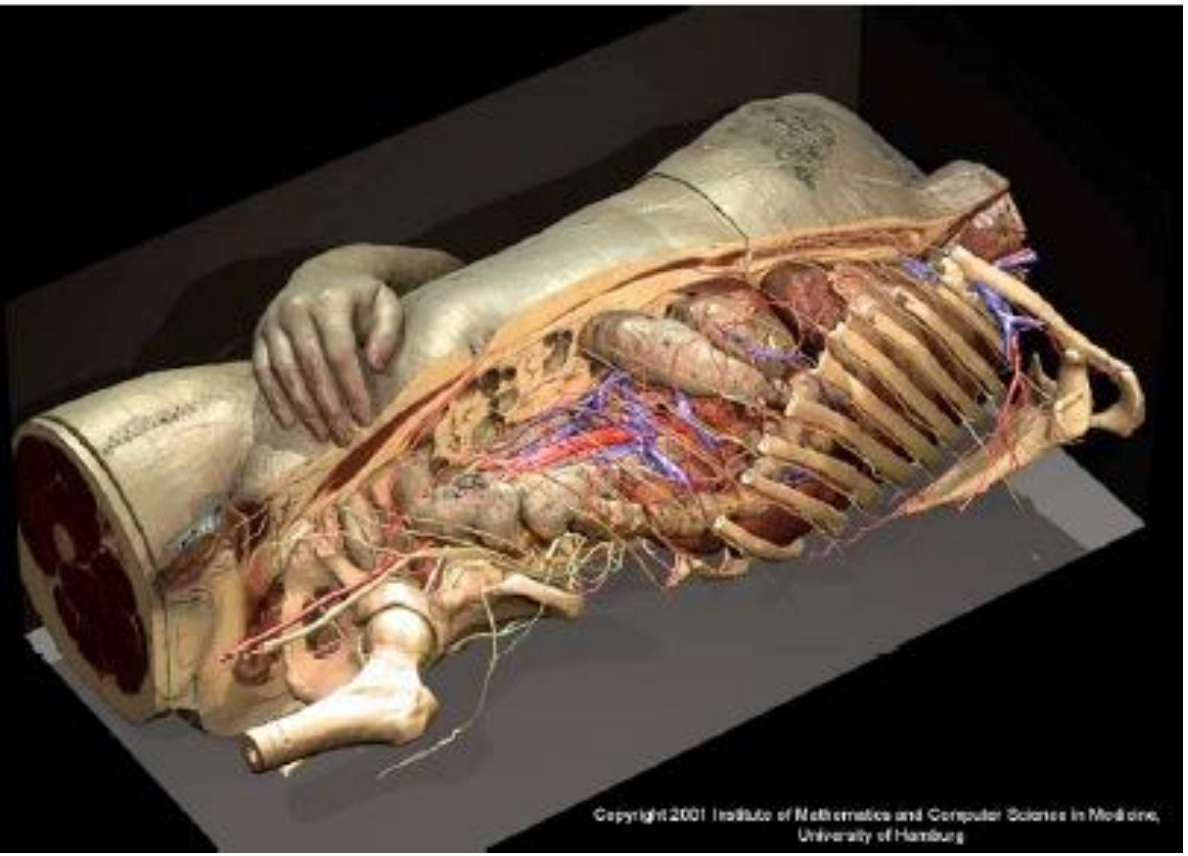
VISUALIZATION AREAS. MEDVIS

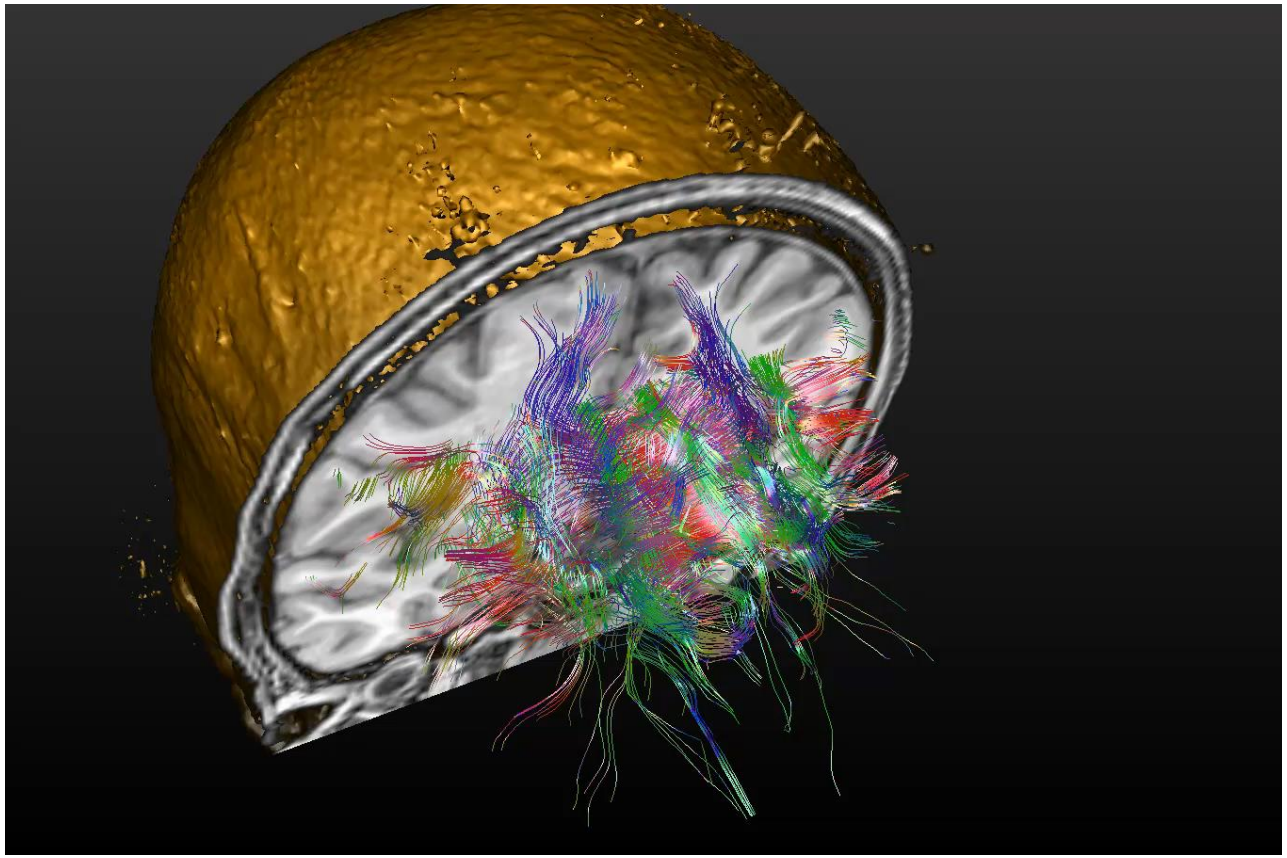


VISUALIZATION AREAS. MEDVIS



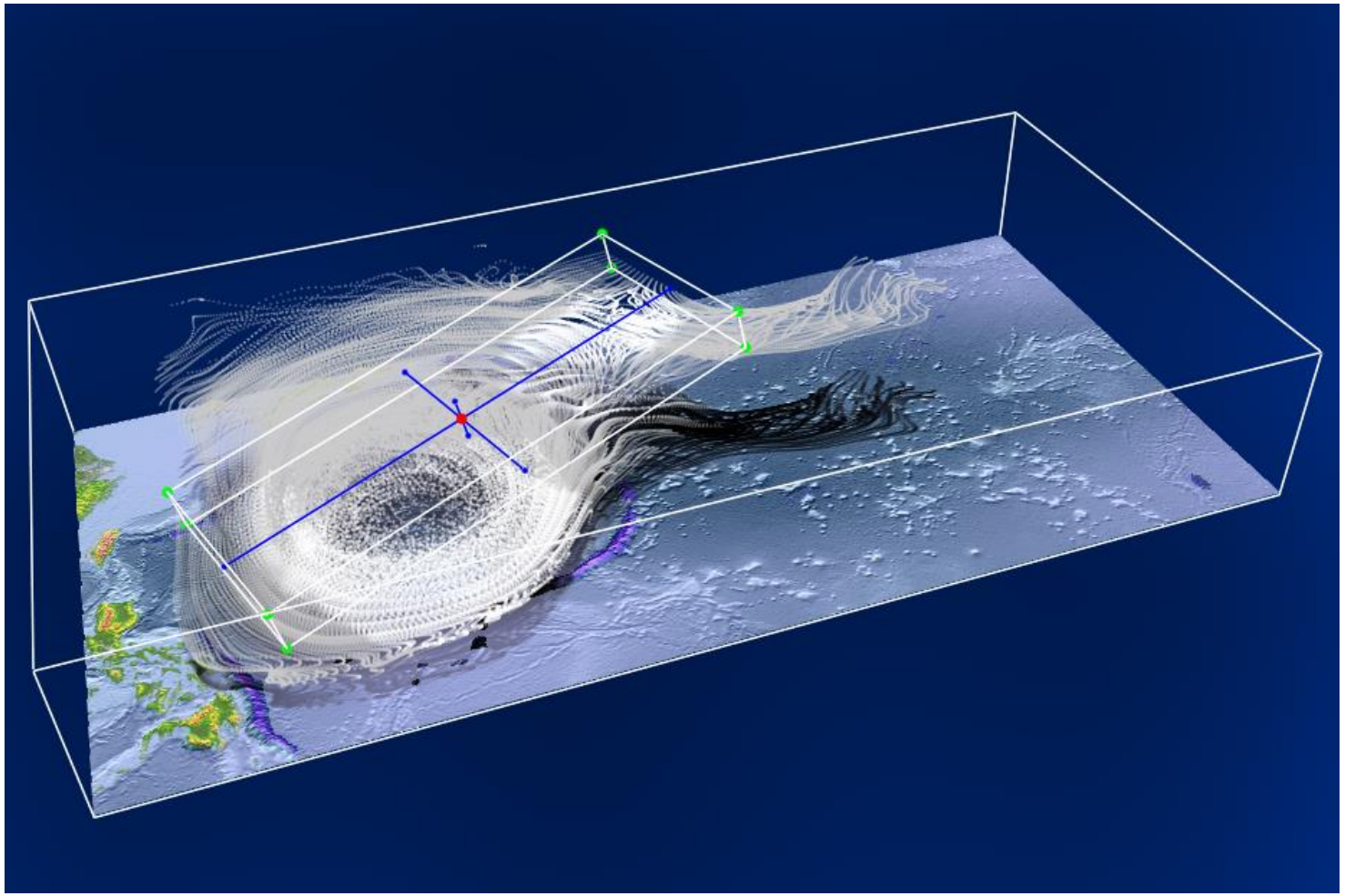
VISUALIZATION AREAS. MEDVIS



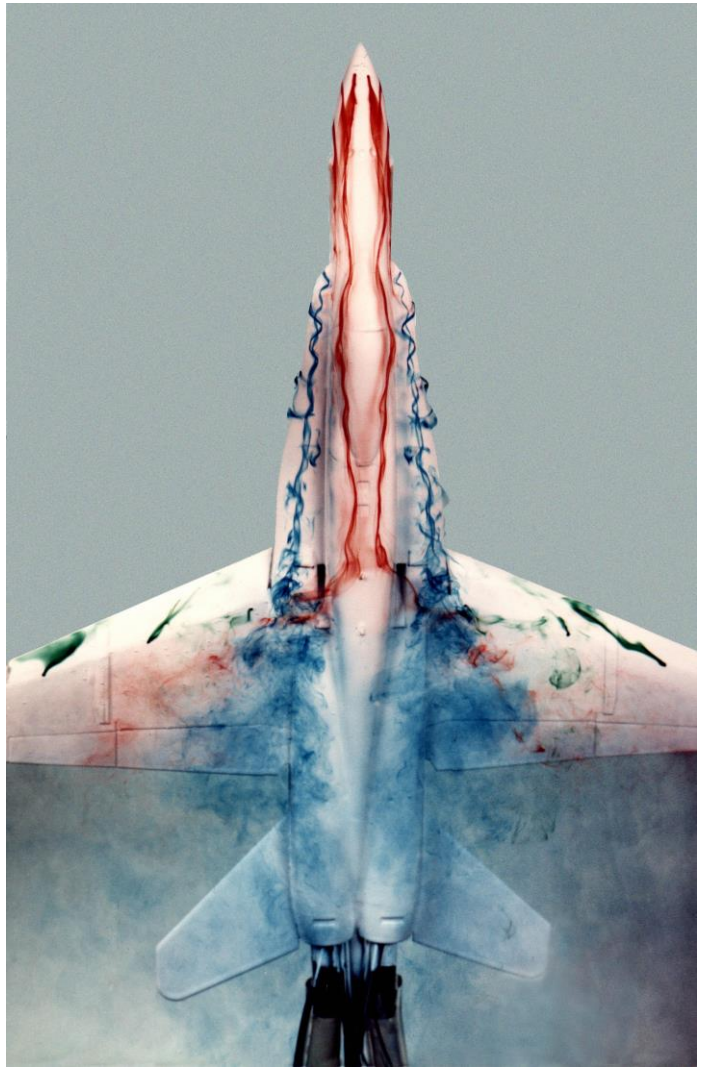


Brain white matter visualization

VISUALIZATION AREAS. FLOWVIS

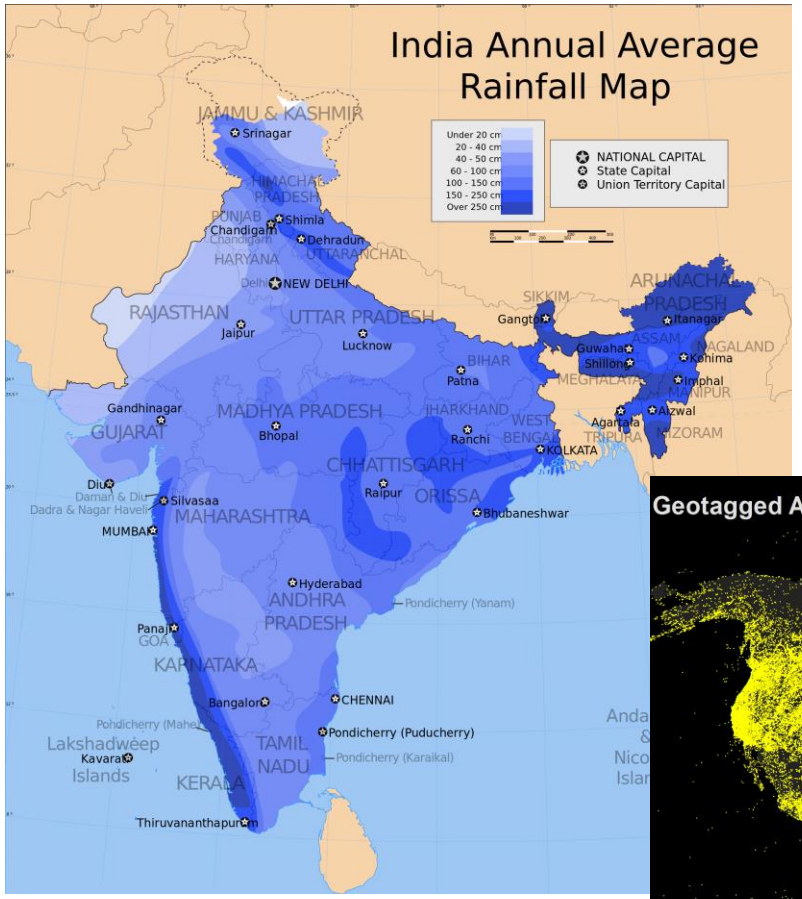


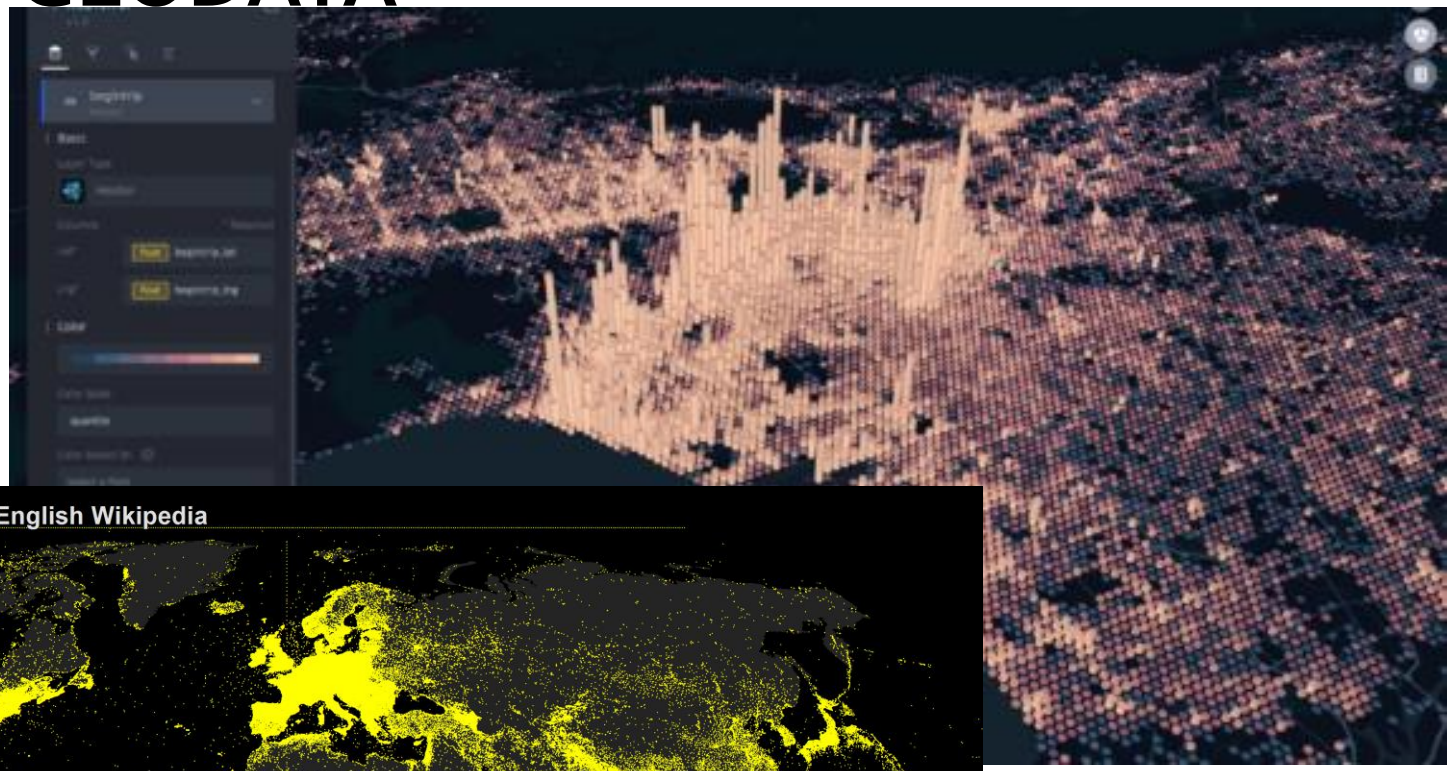
atmospheric physics,
meteorology

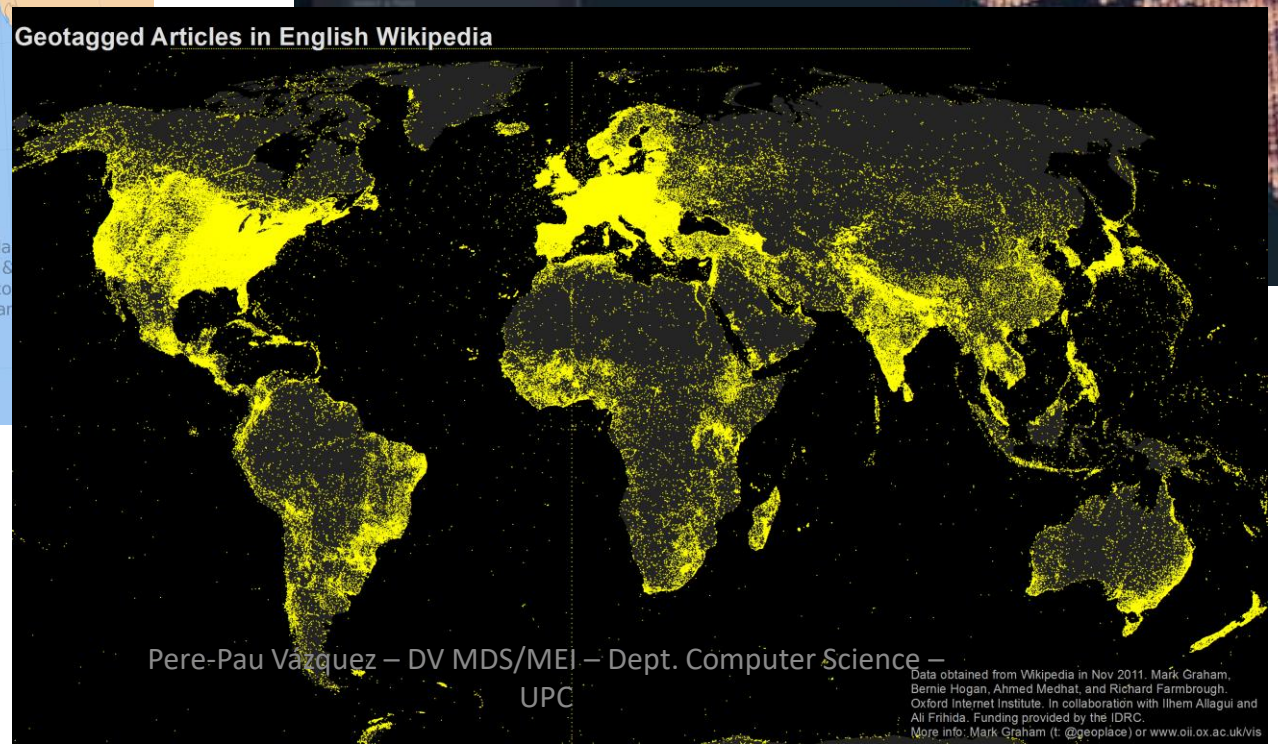


Aircraft design

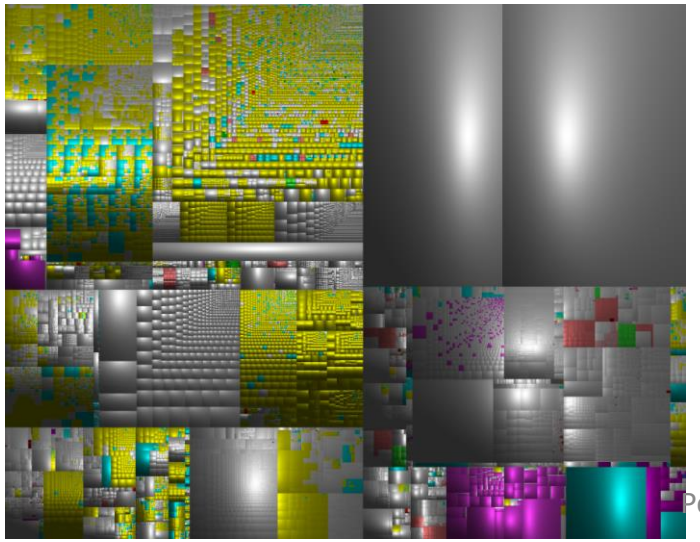
VISUALIZATION AREAS. GEODATA

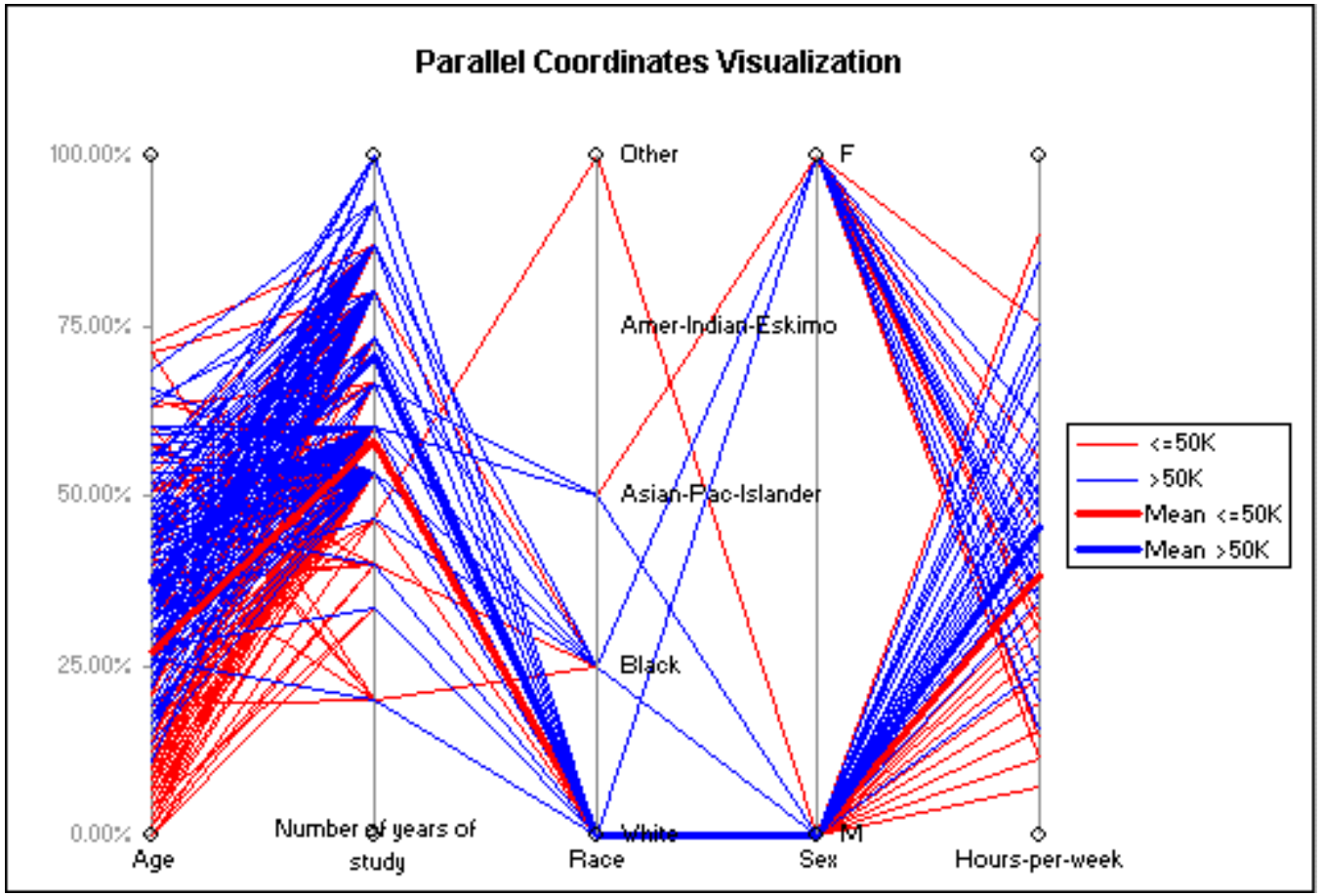




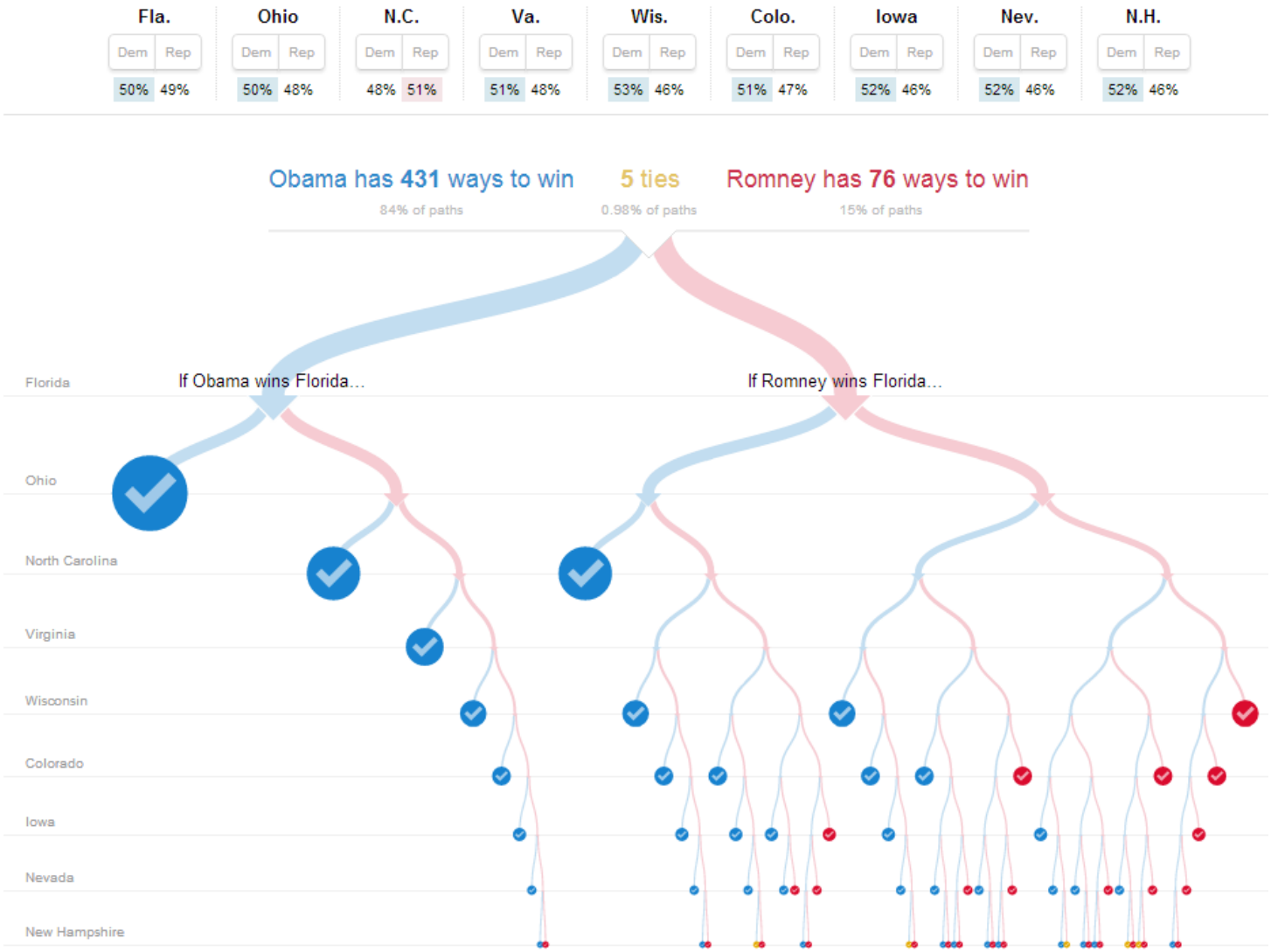


VISUALIZATION AREAS. INFOVIS



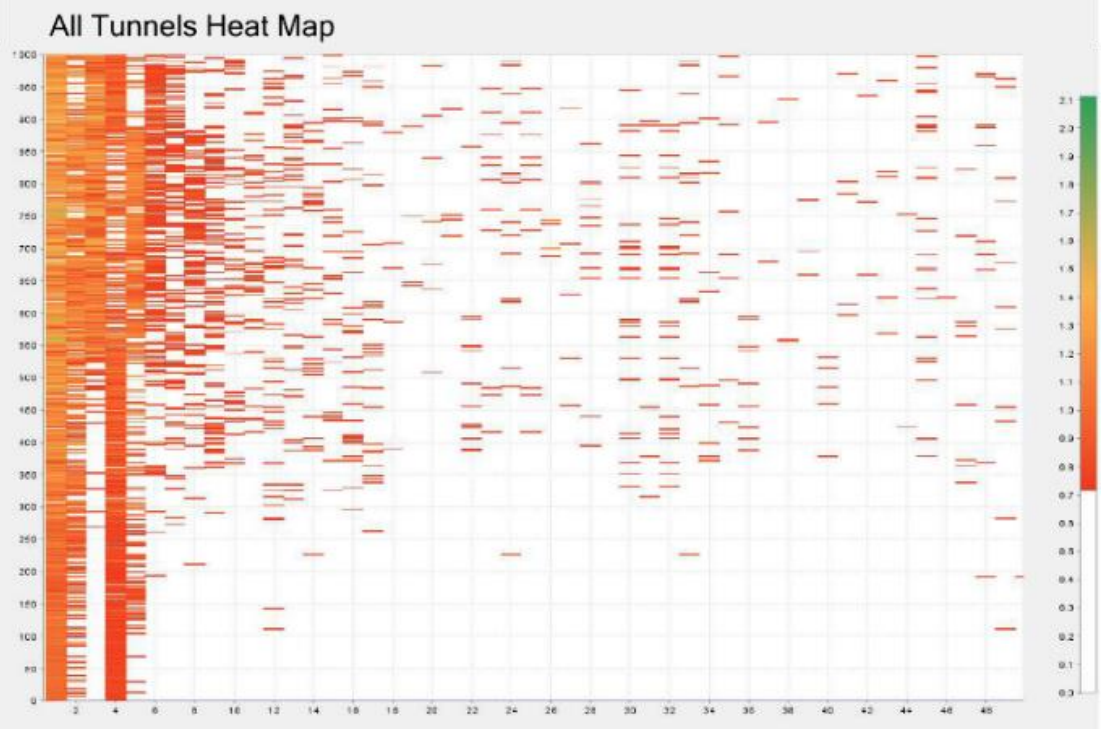


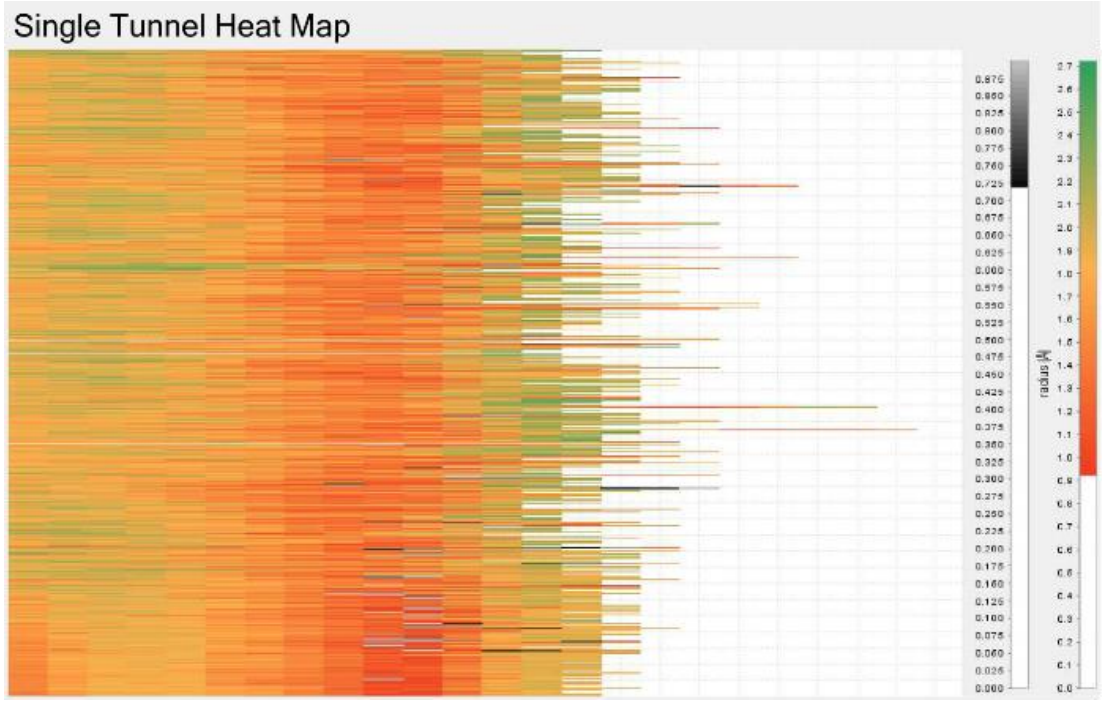
VISUALIZATION AREAS. INFOVIS



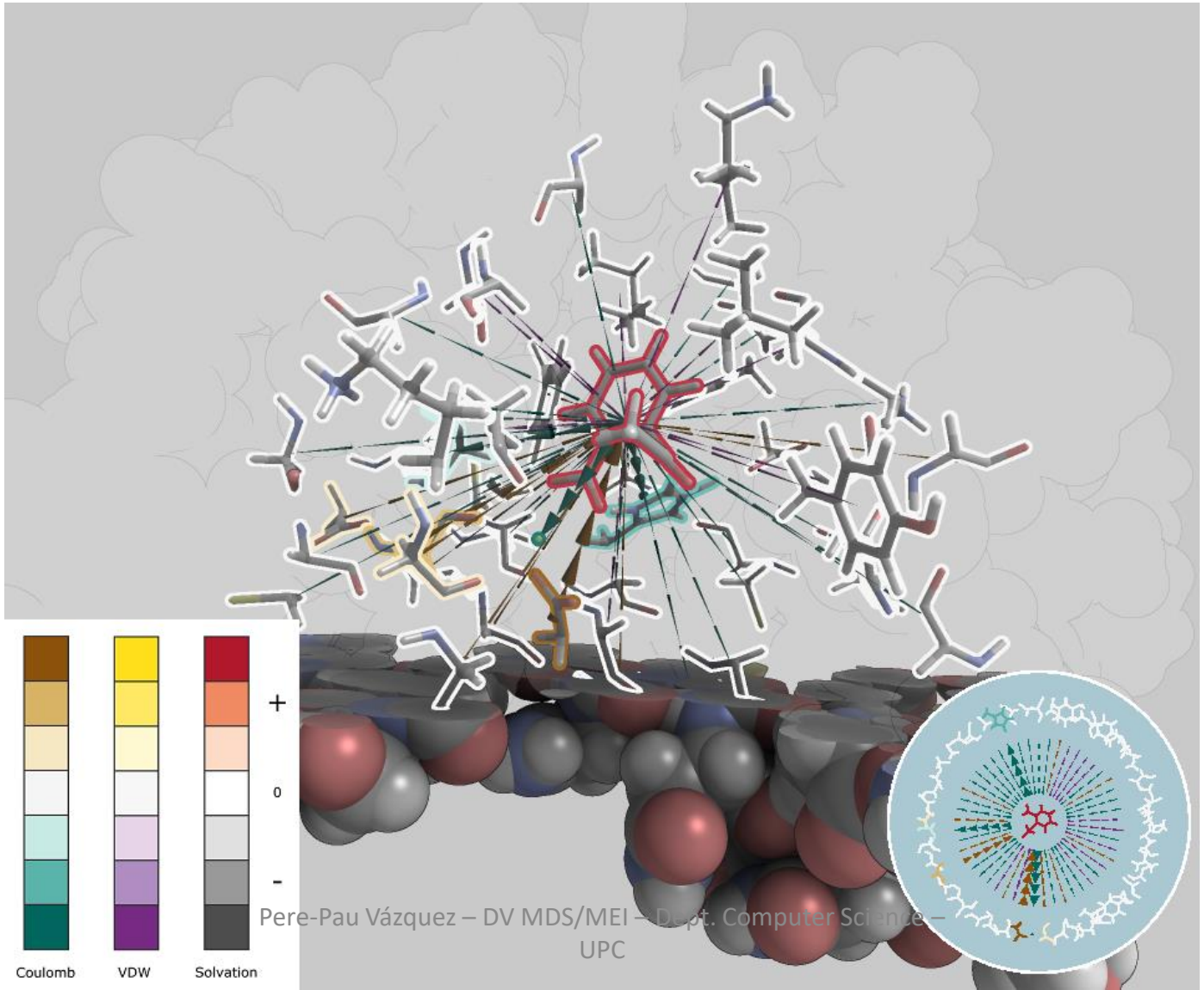
http://elections.nytimes.com/2012/results/president/scenarios

VISUALIZATION AREAS. MOLVIS

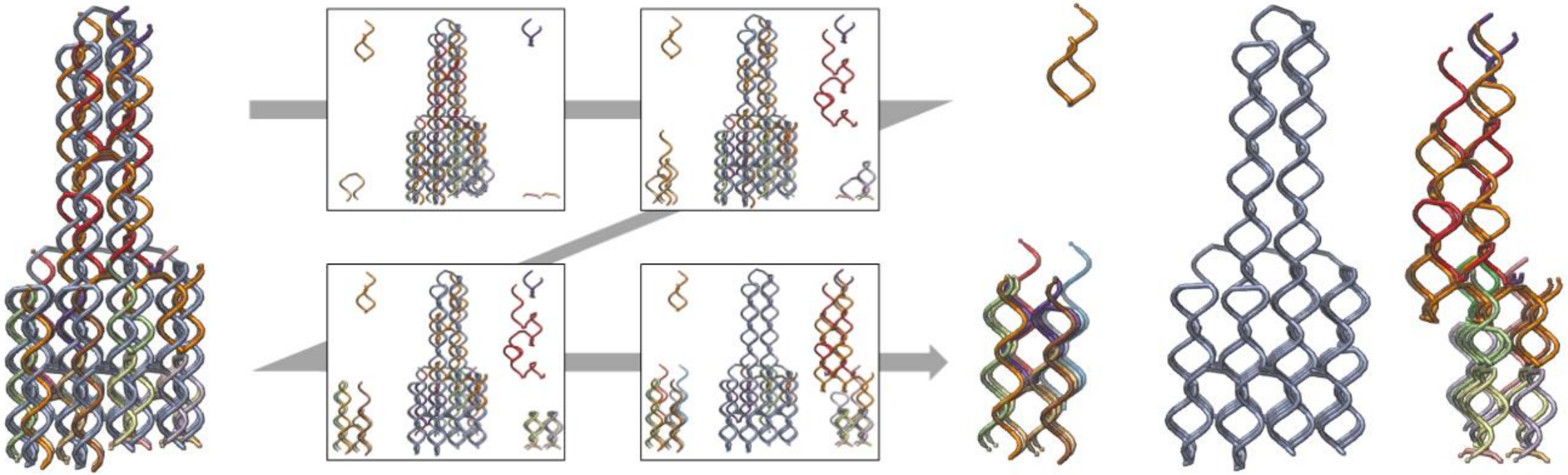




VISUALIZATION AREAS. MOLVIS

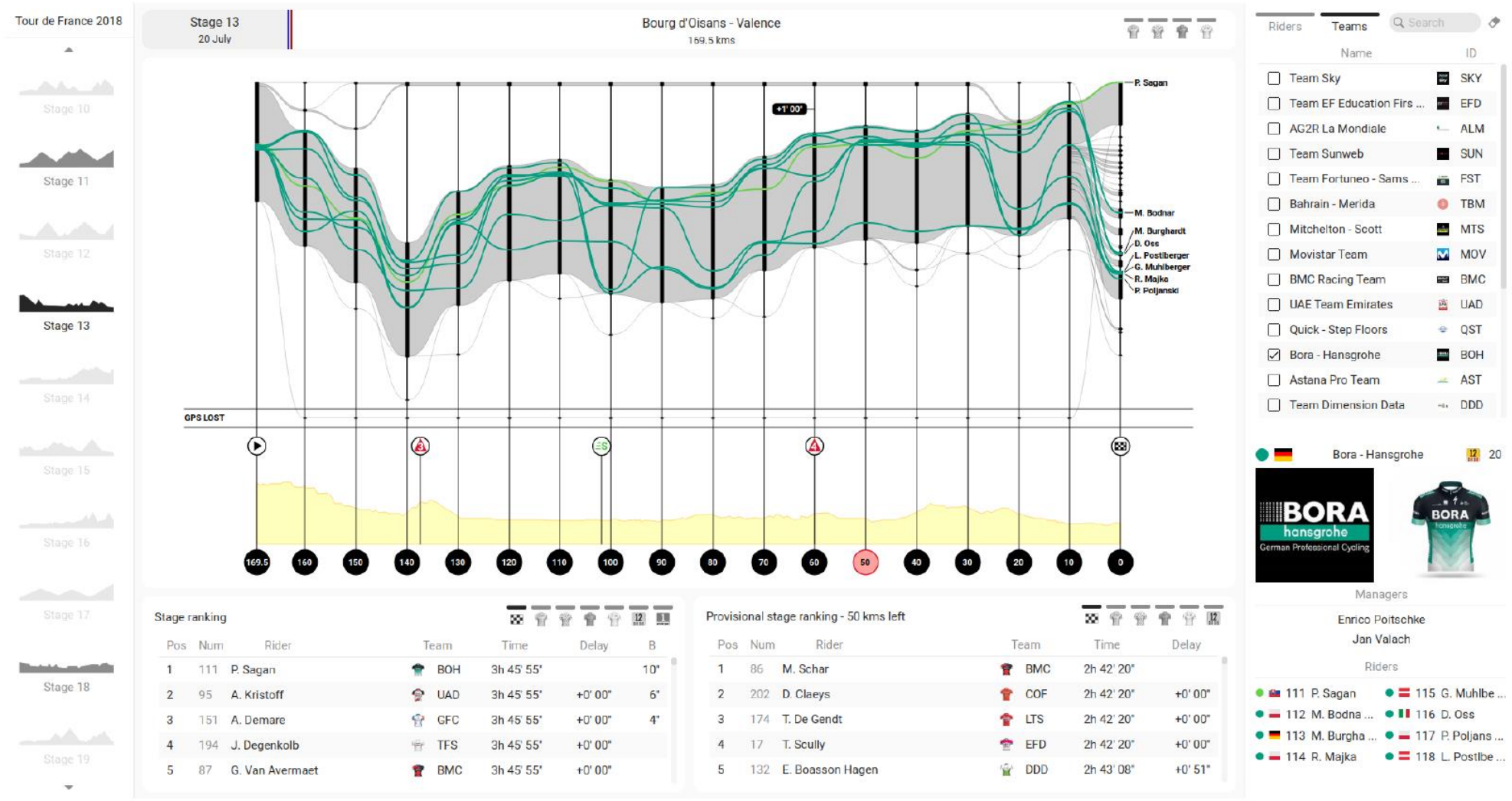


VISUALIZATION AREAS. BIOLOGY



Sbardellati et al., 2019, Interactive Exploded Views for Molecular Structures

VISUALIZATION AREAS. SPORTS ANALYTICS



Díaz et al., 2021, TourVis: Narrative Visualization of Multi-Stage Bicycle Races

VISUALIZATION AREAS. SCIVIS VS INFOVIS VS VA

InfoVis	SciVis (MedVis/FlowVis/BioVis/MolVis...)
Abstract Data No spatial reference.	Spatial Data Spatial reference.
N-dimensional. Heterogeneous.	Mostly 2 or 3-dimensional
Numerical, text, images, multimedia.	Scientific, engineering, biomedical.
Visual Analytics	
Spatial and no spatial data.	
N-dimensional space – hetorogeneous	
Combination with pattern recognition, statistical data analysis machine learning…	

VISUALIZATION AREAS

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC