
Data Science for CSOs

Nino Macharashvili
ForSet

What we are going to talk about today

Let's define the terms

Case studies of Data Science use for social good

How to make CSO more data-driven?

How to use data in your communications?

Where to start from?

Let's define the terms

Data Science



Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.

Let's define the terms

Artificial Intelligence (AI)



Artificial Intelligence (AI) refers to machines (systems and devices) being able to perform tasks that ordinarily require human intelligence in a way that we would consider “smart”.

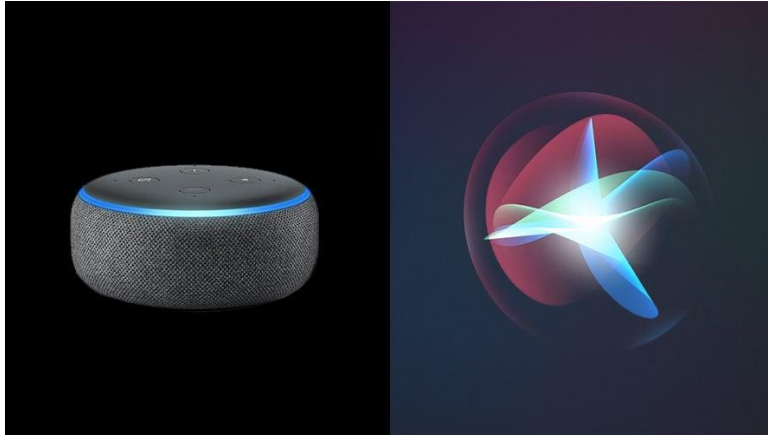
Let's define the terms

Machine Learning

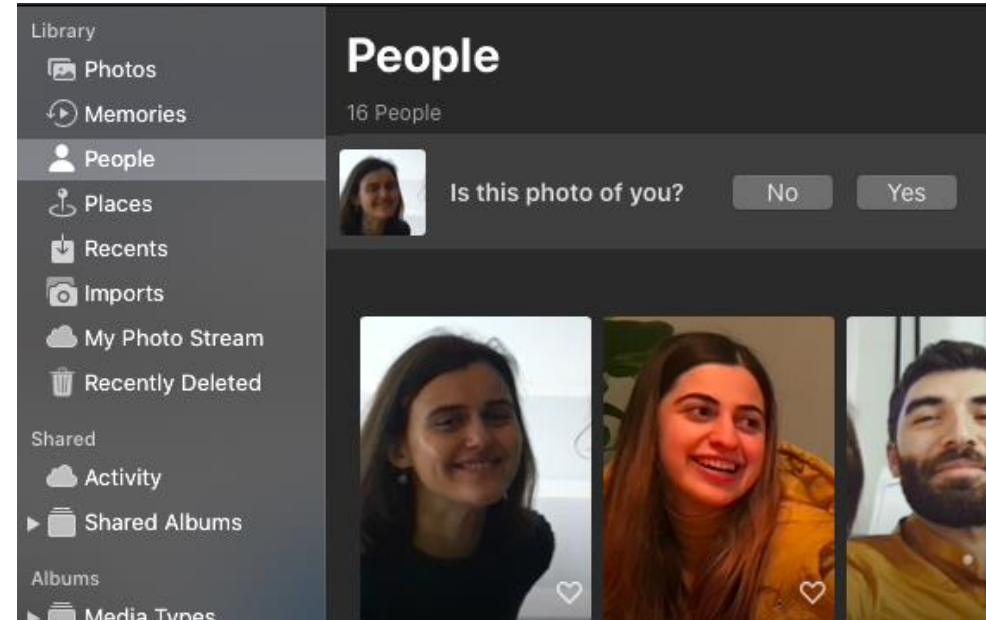
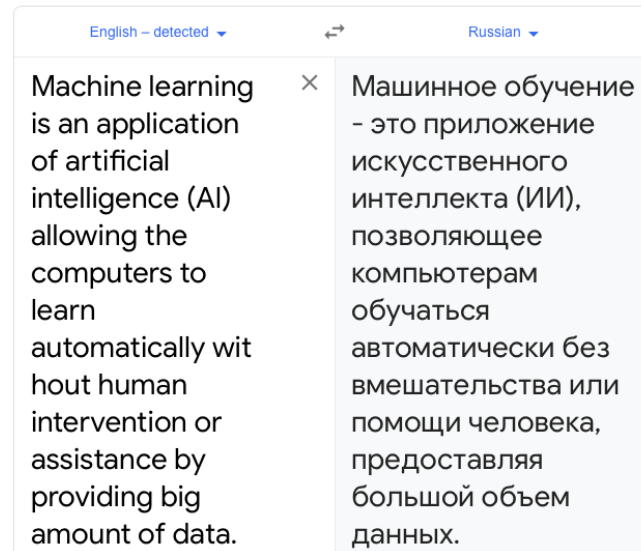
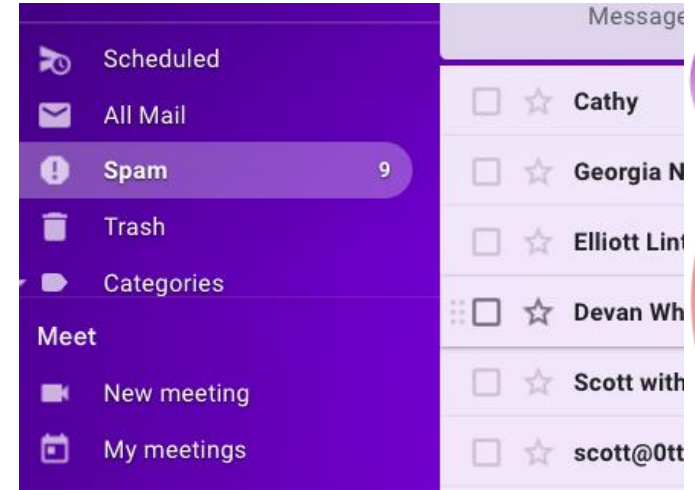


Machine learning is an application of artificial intelligence (AI) allowing the computers to learn automatically without human intervention or assistance by providing big amount of data.

EXAMPLES?



- Personal Assistants (voice recognition)
- Email Spams (text recognition)
- Photos tagging (Image recognition)
- Home appliances
- Automatic Translations



Use of Data Science in CSOs work

Data Science for CSOs

While most of the projects are still under development, the early results show the potential of Data Science applications being widely used in CSOs work and social good broadly, by:

- adapting the technologies developed by commercial sector
- automatizing manual work of monitoring and analysis and instead concentrate on interventions
- do data-driven predictions to save lives

Human Rights

Machine Learning models can assist Human Rights Defenders to analyze and categorize large amounts of data .

Hate Speech

Advanced text analysis can detect and measure hate speech online and empower organizations to focus on interventions to reduce hate speech over time.

Elections

By knowing how political parties are using big data and data science in their campaigns to sway elections, CSOs can watchdog more effectively.

Disinformation

From fake news to fake videos, AI is widely used to spread disinformation. The problems created by technology can only be tackled with the same technology.

Environment

From forecasting the climate more accurately to developing solutions reserving the resources, data science is widely used to protect environment.

Use of Data Science for Human Rights

Machine learning tools help human rights practitioners to decrease the amount of time needed to categorize and analyze the evidence by:

- Classifying sentences and adapting to the specific research questions human rights defenders are interested in
- Running video analysis detecting objects, sound, speech, text and event type to document the war crimes.

- Benetech, in collaboration with the UN and Syrian civil society organizations is developing an AI model to use the estimated 4 million videos related to the conflict as evidence in order to assess human rights violations and promote accountability and the rule of law in Syria and conflict settings worldwide.

<https://benetech.org/lab/ethical-ai-to-promote-justice/>

- Another initiative led by the Swansea University is developing machine learning algorithms to identify and analyze UK-manufactured cluster munitions in crowdsourced footage of more than 20 thousand airstrikes in Yemen's civil war from 2015.

<https://www.technologyreview.com/2020/06/25/1004466/ai-could-help-human-rights-activists-prove-war-crimes/>

Use of Data Science in Detecting Hate Speech

According to the Pew Research Center, **41% of American adults have experienced online harassment, and 66% have witnessed it.**

Using the Natural language processing (NLP) and machine learning (ML) tools, a joint initiative of UC Berkeley's D-Lab and the Anti-Defamation League, and in collaboration with Reddit, Google, and Facebook, is creating **a scalable tool that can be deployed on internet content to**

discover the scope and spread of online hate speech. This tool will uncover trends across different online platforms, allowing the advocates to push for the technical, legal, social, and policy changes necessary to ensure that online communities are safe and inclusive spaces.

More about the project: <https://www.adl.org/resources/reports/the-online-hate-index>

Workshop recording on basic machine learning concepts for HRDers: <https://youtu.be/ZQol4sVfB-U>

Use of Data Science in safeguarding Elections

The methods that political parties use data science to win the elections:

- **Digital Listening**

Monitoring and analysing what someone does or says on social media and measuring which topics are being discussed among users at a given time and whether people feel positively or negatively towards a candidate. Using AI, they can carry out text and behaviour analysis on social media to understand how well any company, campaign or individual is viewed by the public, which can be helpful for politicians to understand how they can improve their status in the eyes of potential voters.

- **Psychometric Profiling**

The process by which your observed or self-reported actions are used to define your personality traits, in order to understand and influence individuals' beliefs, behaviours and motivations. By mining vast quantities of personal data, political strategists can tailor their communications to have greater influence on political opinions and voter preferences.

Learn more:

Tactical Tech: Data and Politics: <https://ourdataourselves.tacticaltech.org/projects/data-and-politics/>

Varoon Bashyakarla - The Democratic Consequences of Data-Driven Politics (https://youtu.be/14UupSgz_0s)

Tetyana Bohdanova - Digital Democracy or Data Exploitation?! (<https://youtu.be/9jvPliOLhyU>)

Full report by Tetyana: <https://ourdataourselves.tacticaltech.org/posts/overview-ukraine/>

Use of Data Science in fighting Disinformation



Grover – a model, robot which generates fake news

Human beings believe in Grover-generated news more than the news written by human beings

That's why Grover is better at detecting fakes news than human beings

Project website: <https://rowanzellers.com/grover/#>

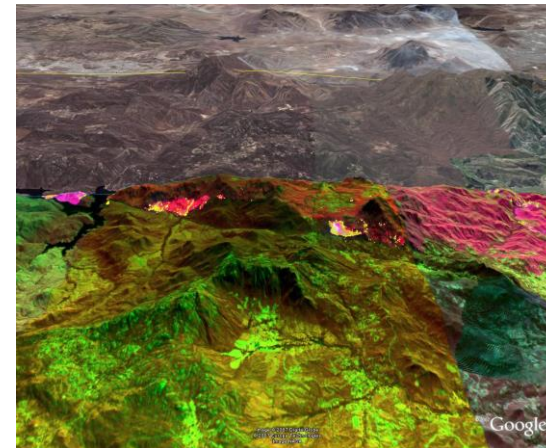
Talk by Camille van Hoffelen, "Grover vs. the Fake News Robot Uprising"

<https://youtu.be/vLYjhJOEcqA>

Use of Data Science to protect environment

Data Science is used in diverse ways for protecting environment, such as:

- Using image recognition to identify endangered species by their unique looks, observe and protect them.
- Predicting forest wildfires by combining historical data with real-time satellite imagery and simulations, mostly using drones.



Remaining challenges and concerns

How accessible is Data Science & AI for CSOs?

- Data Science is an expensive skill-set, often not affordable for CSOs
- Access to data needed to train models is still limited for CSOs
- There is a remaining gap between the private sector and its innovation and CSOs/non-profits
- Developing working models take time. CSOs usually do not have ability to do long-term planning and investing

What are the limitations of Data Science for CSOs work?

- Bias and fairness
- Privacy
- Safe use and security
- “Explainability” (the ability to identify the feature or data set that leads to a particular decision or prediction).

Potential Solutions

Tap into available resources

- ✓ Use open-source ready-made applications
- ✓ Engage in programs that can pair you up with volunteer data scientists

Build your and your CSOs' skills and network

- ✓ Use online study materials
- ✓ Find the local community groups
- ✓ Work on real-life data projects and get experience

Change Attitude and become more data-driven

- ✓ Start small
- ✓ Think how Data Science could make your work more efficient – just because it is a buzzword does not mean it is applicable to your work
- ✓ Discuss your ideas with your team and community
- ✓ Follow trends and recent technological developments

Questions?

How to make CSOs more data-driven?

How to start small

Your CSO might not be ready yet to start using the advanced data technologies, such as AI and machine learning tools.

However, you can start small by utilizing the potential of data in your own decision-making, advocacy and communications work.

Record data

AI projects need a lot of data to develop predictive models. Make it a habit.

Make your data open

Respecting the privacy, you should make as much data open as possible. Collaboration is the key.

Advocate for open data

It should not be only governments who open up data, but also private companies and CSOs, by giving an example.

Learn from data

Start asking the questions that data can help answer. Find skillset to analyze data to find the answers.

Act on insights

Change your behavior and actions based on the insights that come from data.

How to use data in your communications?

When can you use data in communications?

1. **You have data** and you need to communicate it to a particular audience;
2. You are communicating about a specific issue, product, or event to a particular audience and you need **to make your communications more effective.**



Where can you find data?

1. National statistical organisations
 - <https://www.ons.gov.uk> (UK)
 - <https://www.geostat.ge/en> (Georgia)
2. Local, regional, and national governments and departments
3. International bodies doing global studies
 - World Happiness report (<https://worldhappiness.report/ed/2020/#read>)
 - PISA (Programme for International Student

Assessment) rankings
(<http://www.oecd.org/pisa/>)

4. Private companies and corporations
 - Annual reports and other publications



Where can you find data?

5. Regulators and auditing bodies
6. Charities and non-profit institutions
7. Professional bodies and unions
8. Online communities and groups
9. Open data initiatives in the field

Tips:

- send out Freedom of Information Requests (FOI) much earlier than your project deadline
- use website monitoring tools to get updates, such as <https://visualping.io>



Use Advance Google Search



Advanced Search

Find pages with...

all these words:

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from: to

To do this in the search box.

Type the important words: tri-colour rat terrier

Put exact words in quotes: "rat terrier"

Type OR between all the words you want: miniature OR standard

Put -ro by... Then narrow your results

- language: Find pages in the language that you select.
- region: Find pages published in a particular region.
- last update: Find pages updated within the time that you specify.
- site or domain: Search one site (like wikipedia.org) or limit your results to a domain like .edu, .org or .gov
- terms appearing: Search for terms in the whole page, page title or web address, or links to the page you're looking for.
- SafeSearch: Tell SafeSearch whether to filter sexually explicit content.
- file type: Find pages in the format that you prefer.
- usage rights: Find pages that you are free to use yourself.

https://www.google.com/advanced_search

Advanced Search
Advanced Search

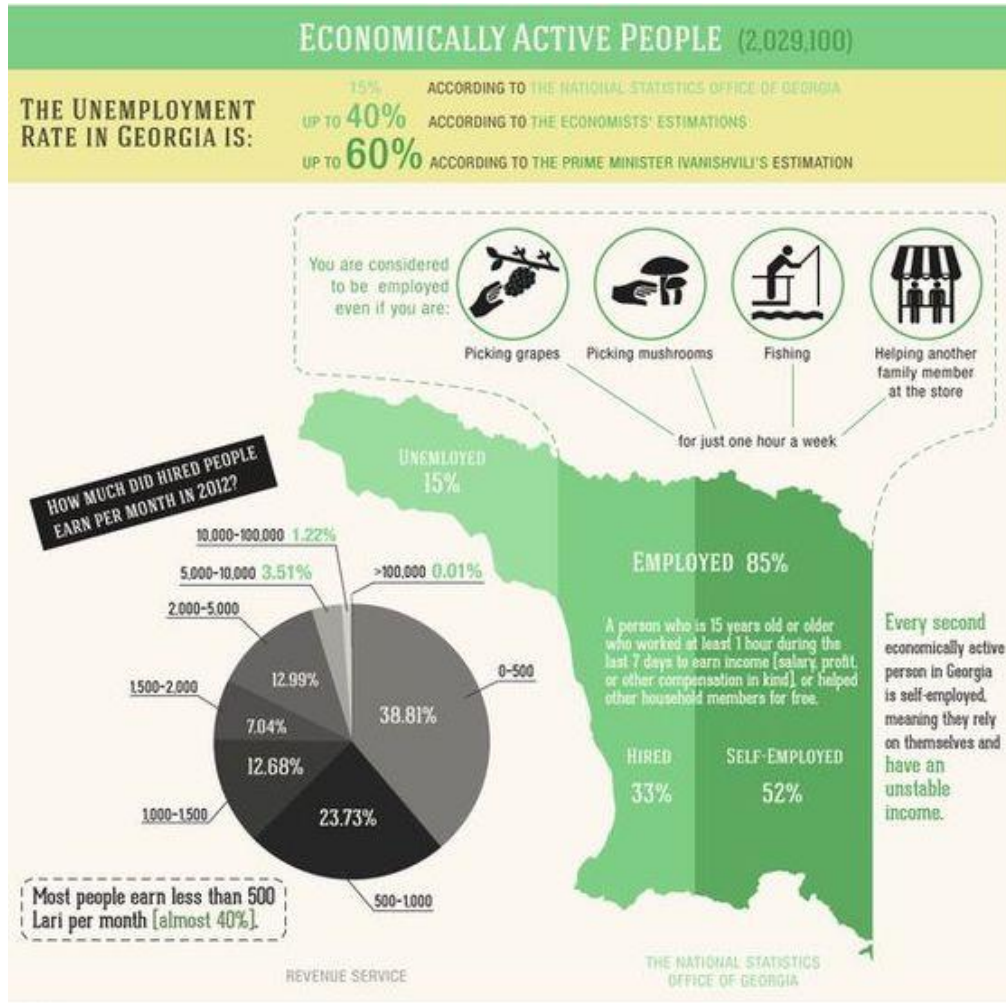
Know your data!

Check accuracy of the data

- Who collected the data?
 - When did they collect it?
 - How did they collect it?
 - What does each term mean?
- *If necessary, find another source of the same sort of data for comparison.*

- *Always include the source name and, if possible, source files (pdf, spreadsheet, link, etc.)*





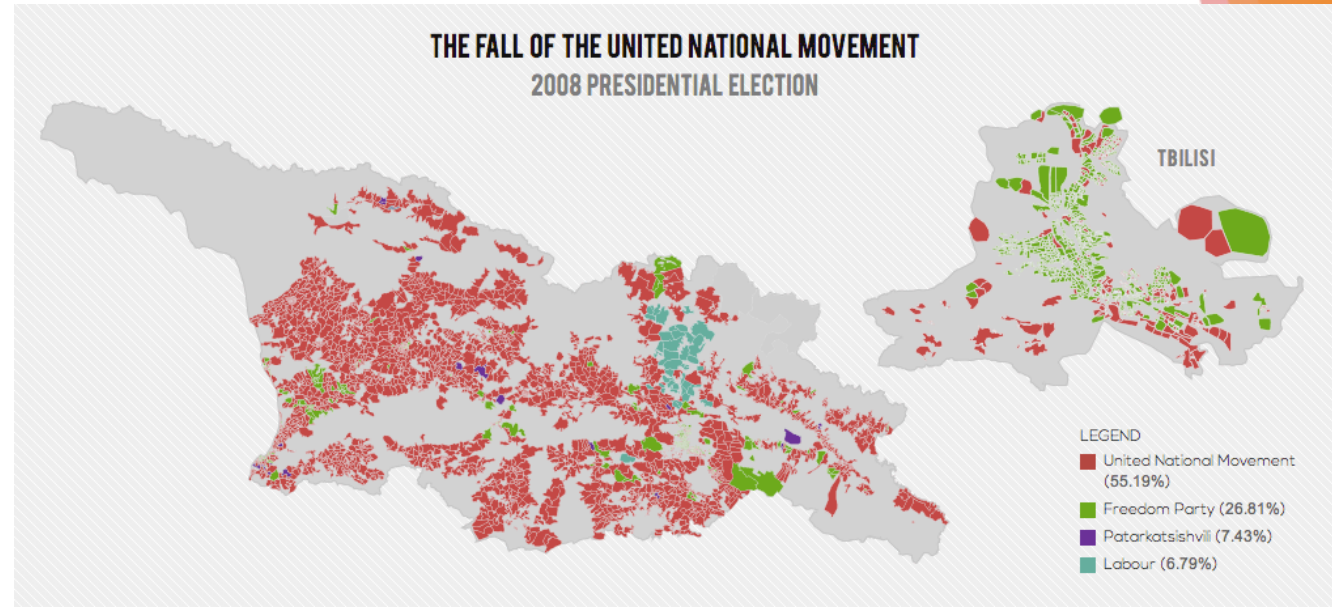
Define the terms

Unemployment is Georgia's N1 problem, JumpStart Georgia, 2013
<https://jsferadi.forset.ge/en/visualizations/unemployment-is-georgians-1-problem>

What is data visualization?

- Graphical representation of Data
- A tool for data analysis and understanding

Often data visualization is the only way to quickly analyze large dataset, understand it and see the trends

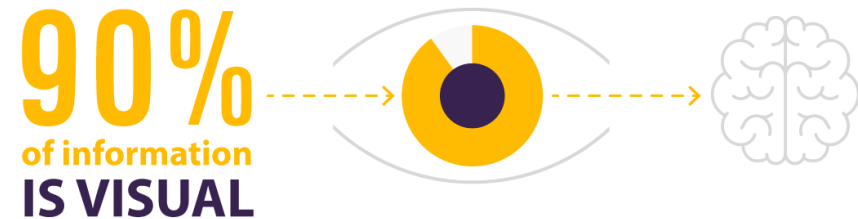


Visualization types

Depending on the data, audience, message, purpose, and distribution channels, your data visualization might get different forms:

- **Factographs**
- **Static infographic**
- **Interactive infographic**
- **Scrolling data story**
- **Data dashboard**
- **Animation**

- **Installation**
- **Sonification**
- **Etc.**



Factographs

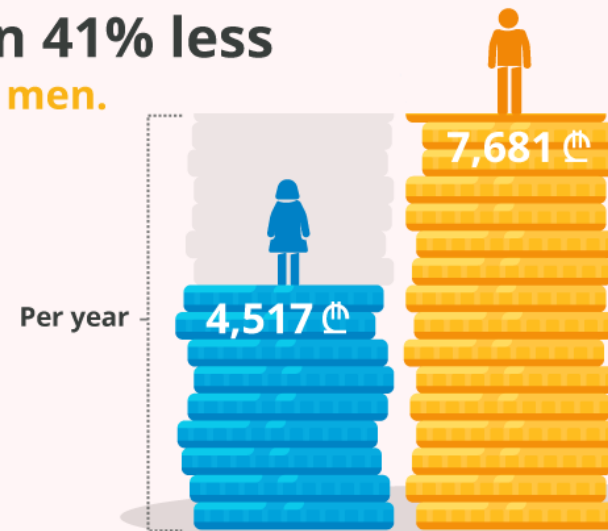
WORLD TB DAY MARCH 24

6 out of 10 children did not access quality care in 2015, risking serious illness and death

World Health Organization

unicef for every child

In Georgia, working women earn 41% less than men.



CONSEQUENCE

A significant number of women report that they do not work, have not sought out a job or left a job due to insufficient pay.

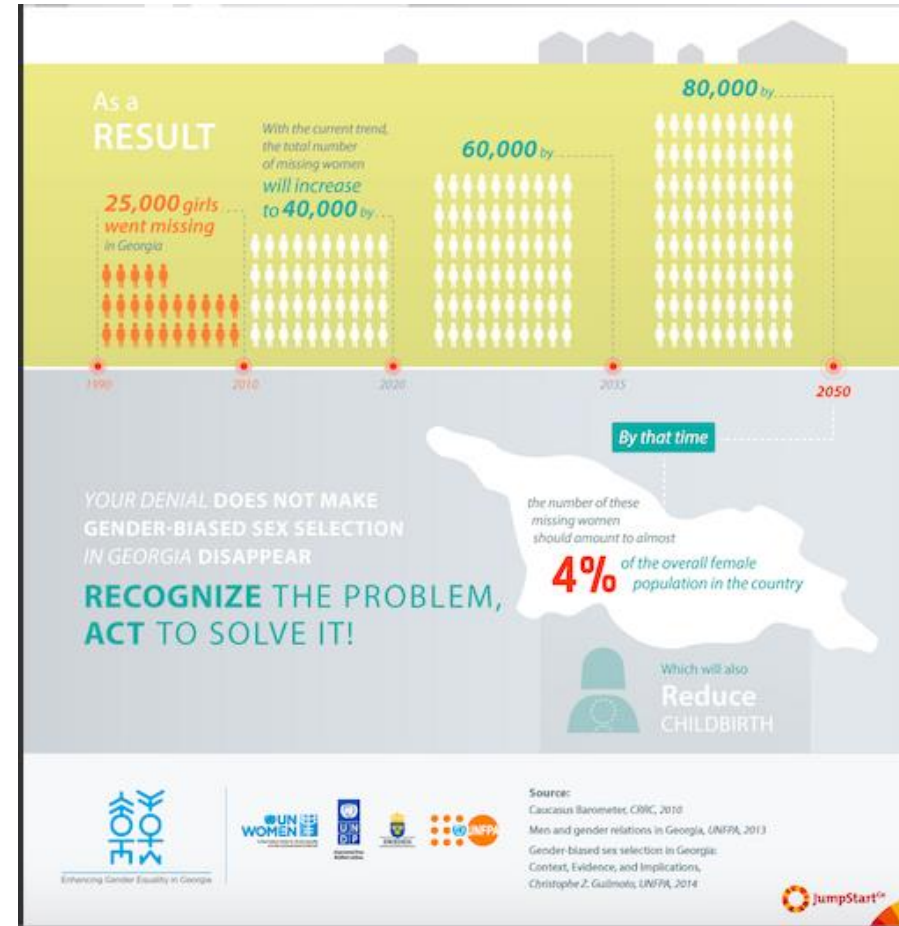
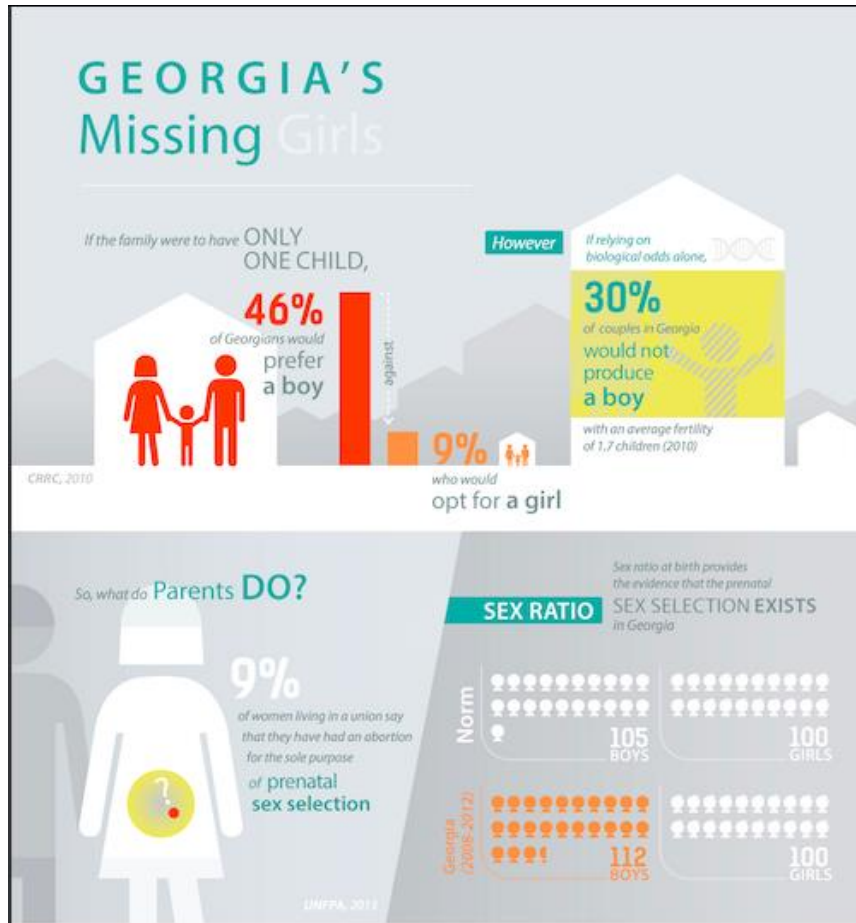
ForSet

Source: The Study "Causes and Consequences of Women's Economic Inactivity and Engagement in the Informal Economy in Georgia", which was conducted in 2018 by CRRC Georgia with the technical support of UN Women and funding from the Swiss Cooperation Office in Georgia.

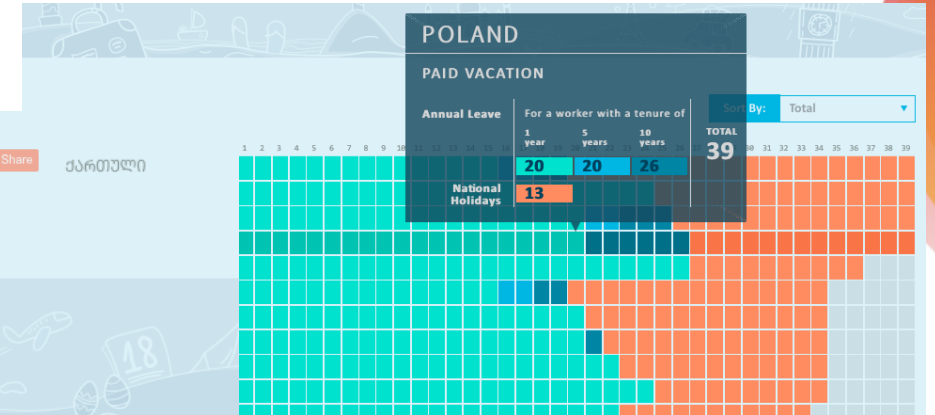
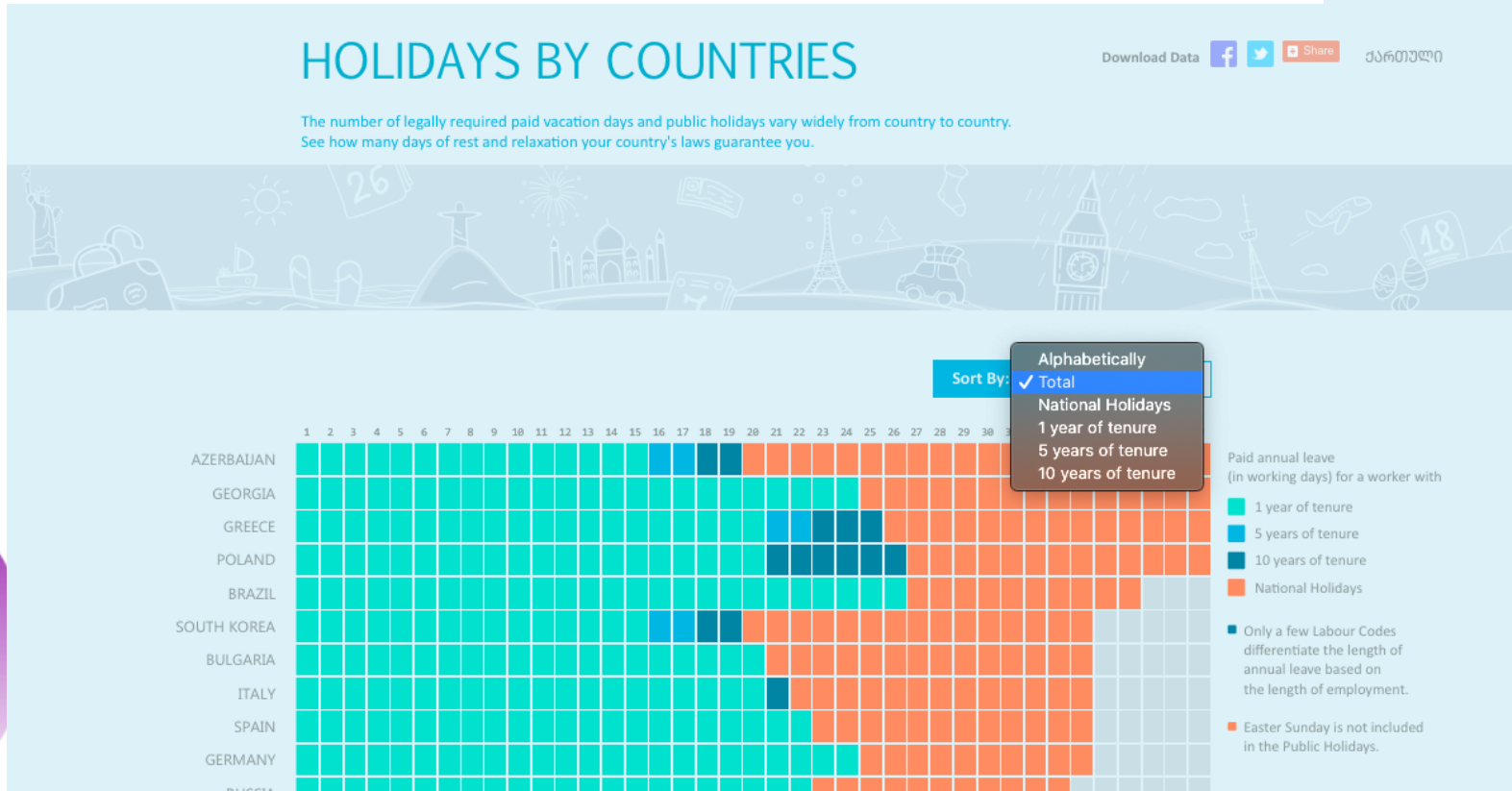


Swiss Agency for Development and Cooperation SDC

Static Infographic



Interactive Infographic



https://jsvisuals.forset.ge/holidays_by_country/en/

Scrolling Data Story



Health

Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve”

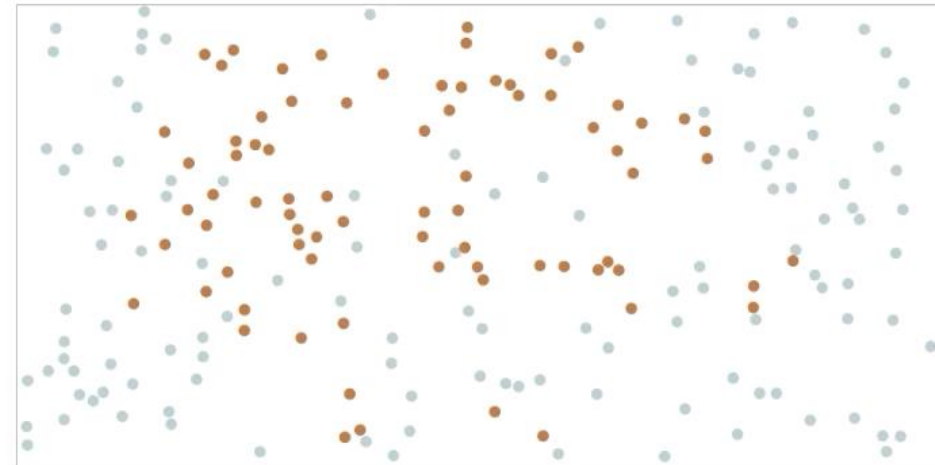
By [Harry Stevens](#) March 14, 2020

<https://www.washingtonpost.com/graphics/2020/world/coronavirus-simulator/>

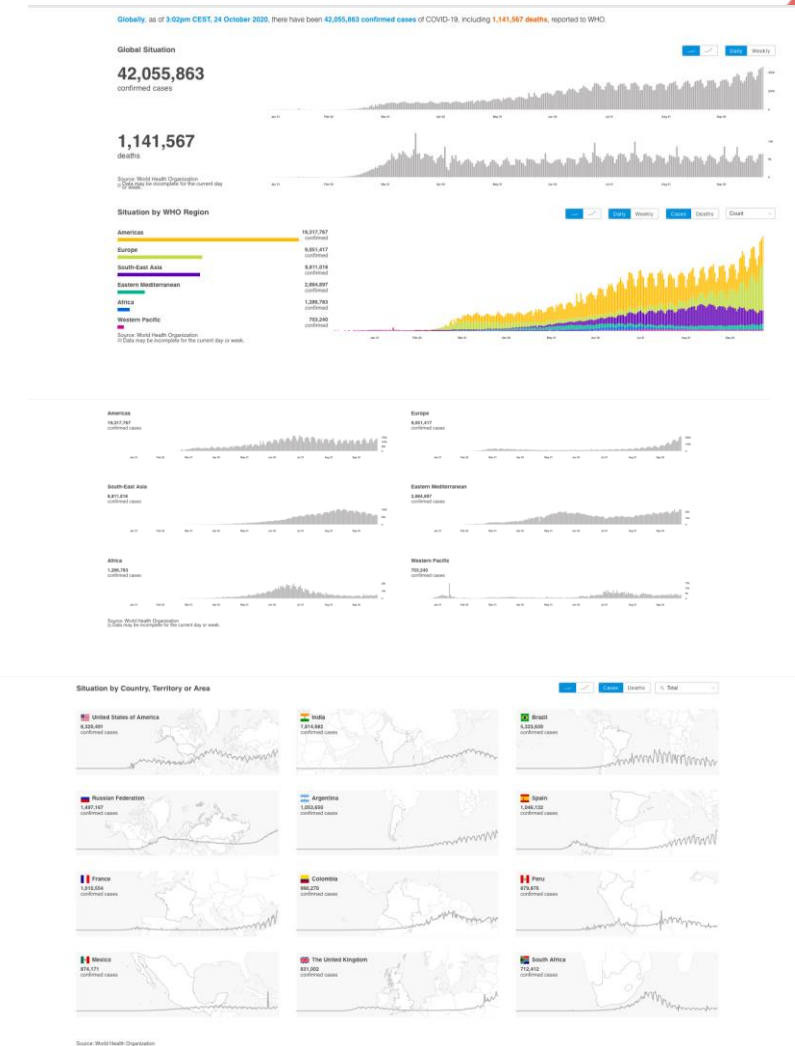
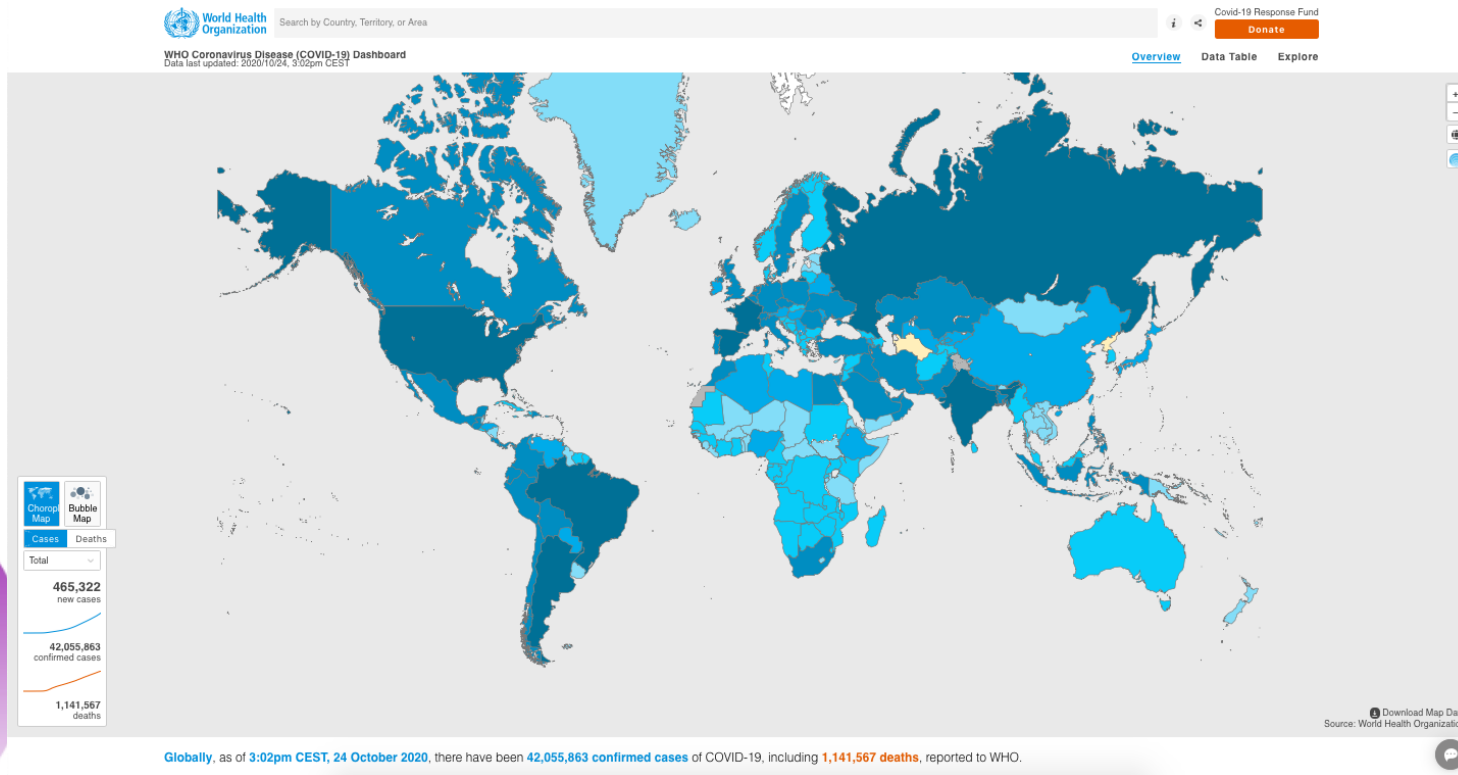
Let's see what happens when similitis spreads in a town of 200 people. We will start everyone in town at a random position, moving at a random angle, and we will make one person **sick**.

Notice how the slope of the red curve, which represents the number of sick people, rises rapidly as the disease spreads and then tapers off as people recover.

Count	Change over time
Recovered	0
Healthy	122
Sick	78



Data Dashboard

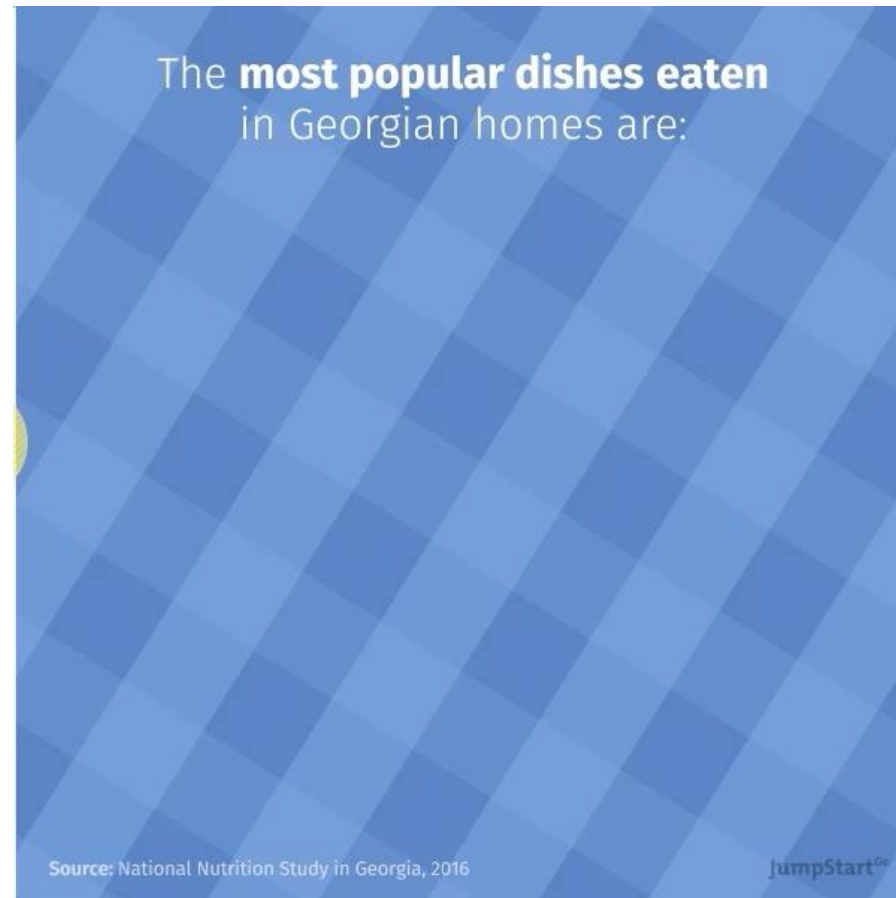


<https://covid19.who.int>

Animation



Gifographic



Creative Data Visualizations



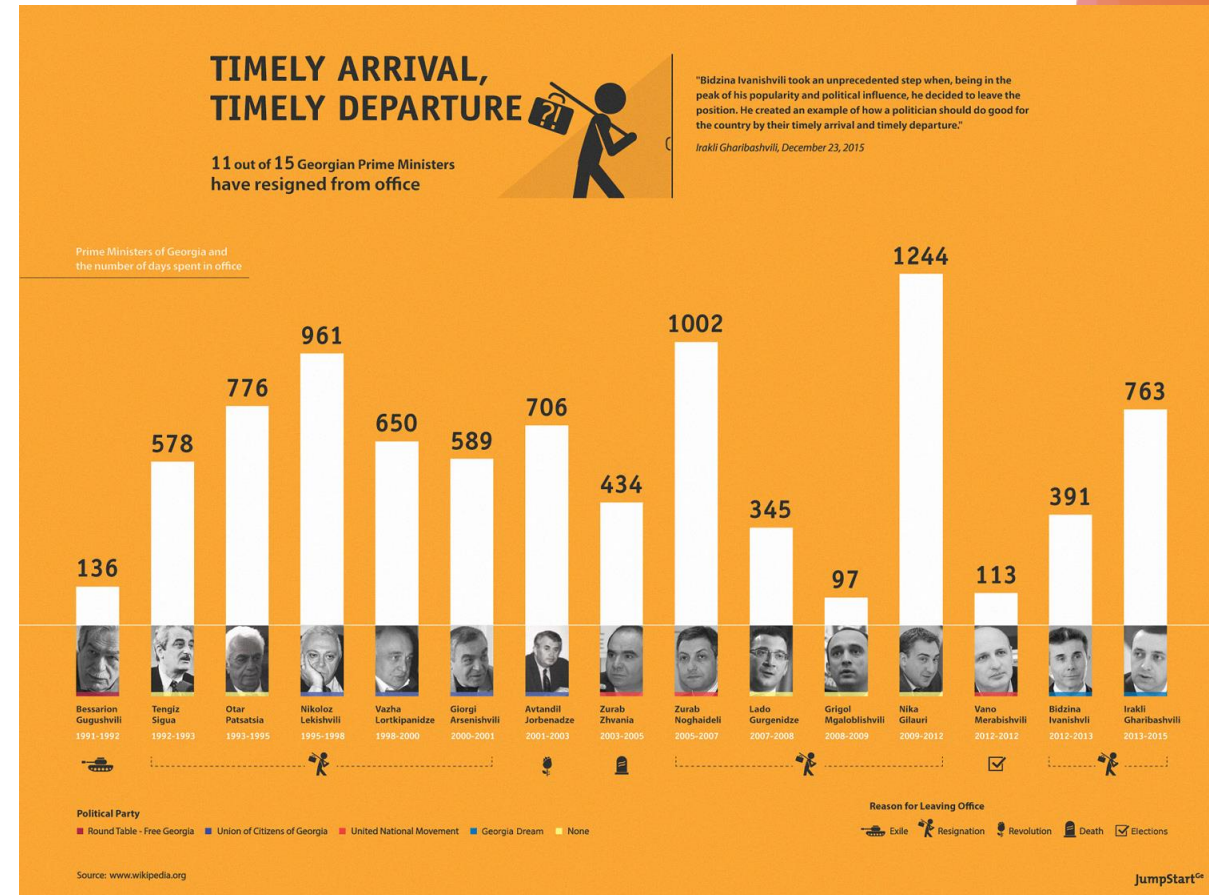
Data-driven album cover by Tiziana Alocci for the Berlin-based record label Sum over histories.
<https://www.tizianaalocci.com/portfolio/raep/>

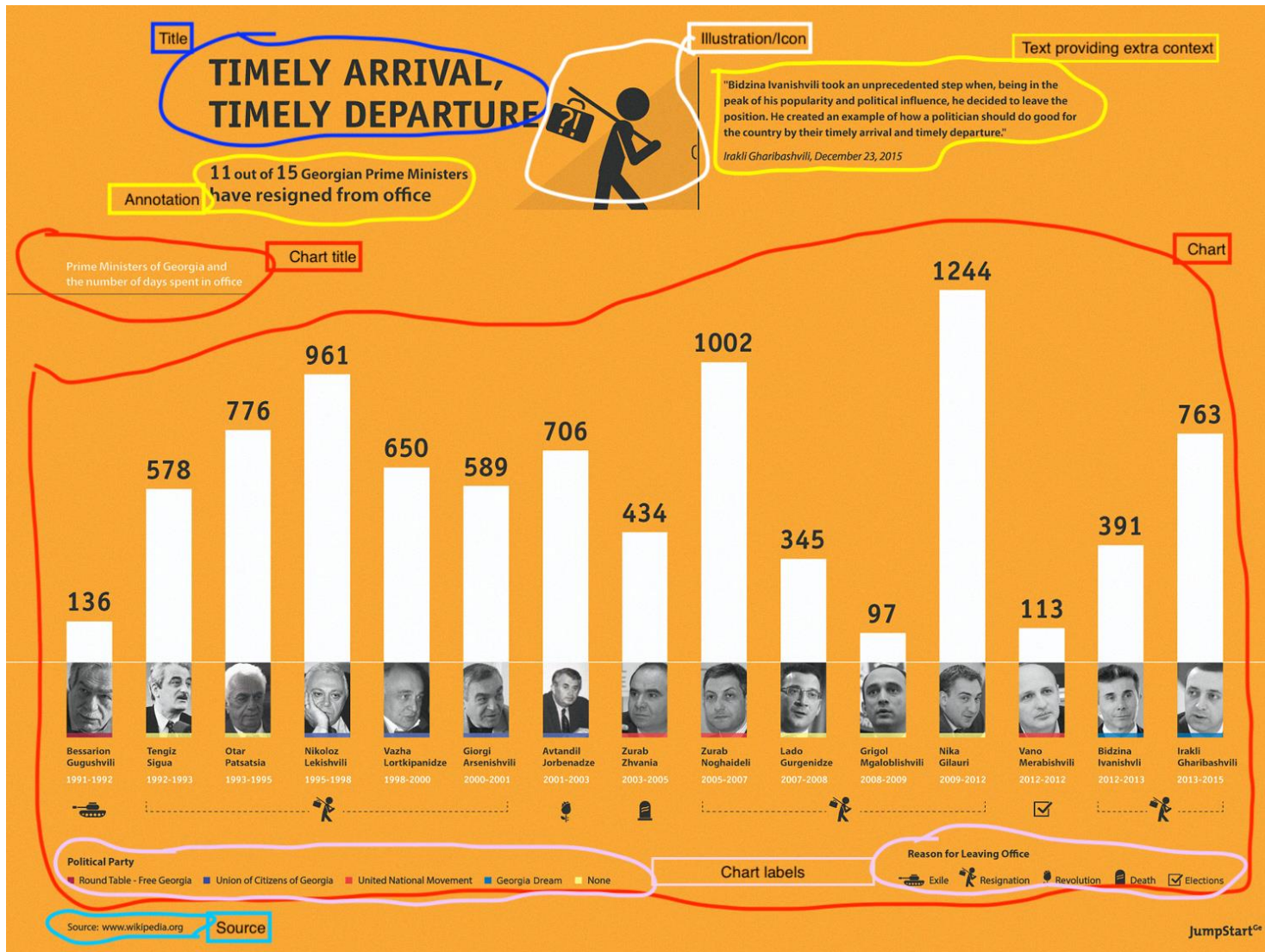
Weather data installation by the artist Nathalie Miebach
https://www.ted.com/talks/nathalie_miebach_art_made_of_storms



Components of data visualization

- **Charts** (with labels, title)
- **Graphic elements** (icons or illustrations, photos)
- **Annotation** (highlights)
- **Text** (title, slogan/call to action, introduction, explanation, mention of data limitations, etc.)
- **Source**





Find the right balance between **the text** and the **visual elements**, to make it easily **understandable and functional**, as well as, **visually appealing**.



If any component does not bring additional value to the visualization, simply remove it.

Kill your darlings!

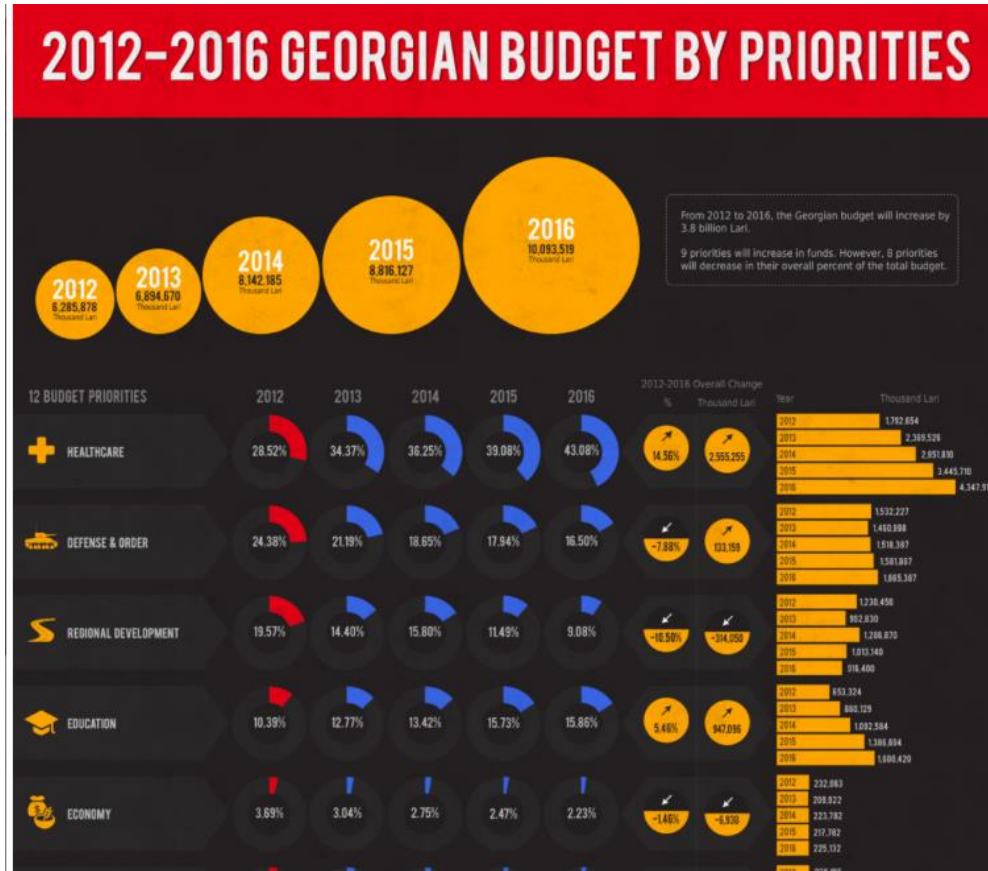
Before choosing the visualization form:

You should answer the following questions:

- Who is my audience?
- What is my message?
- How is my audience going to use this information?
- How am I going to share the visualization with my audience?



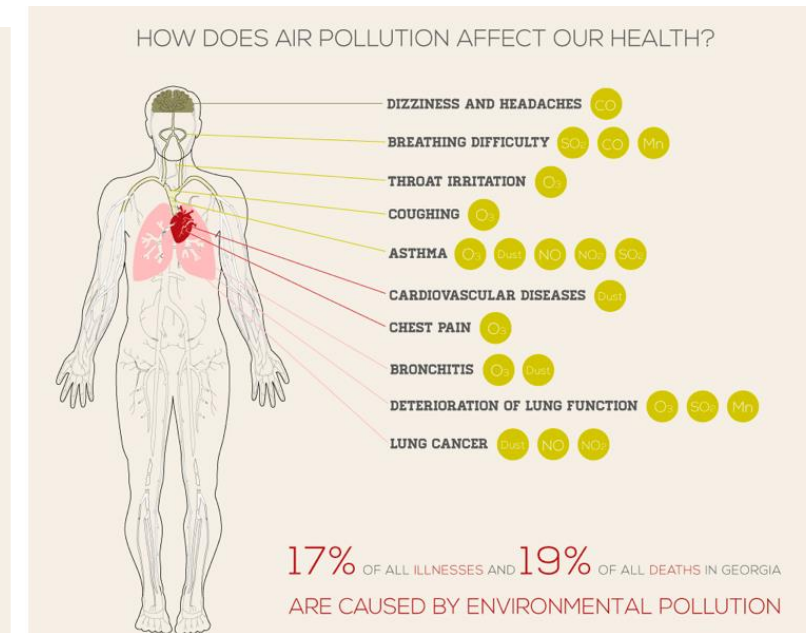
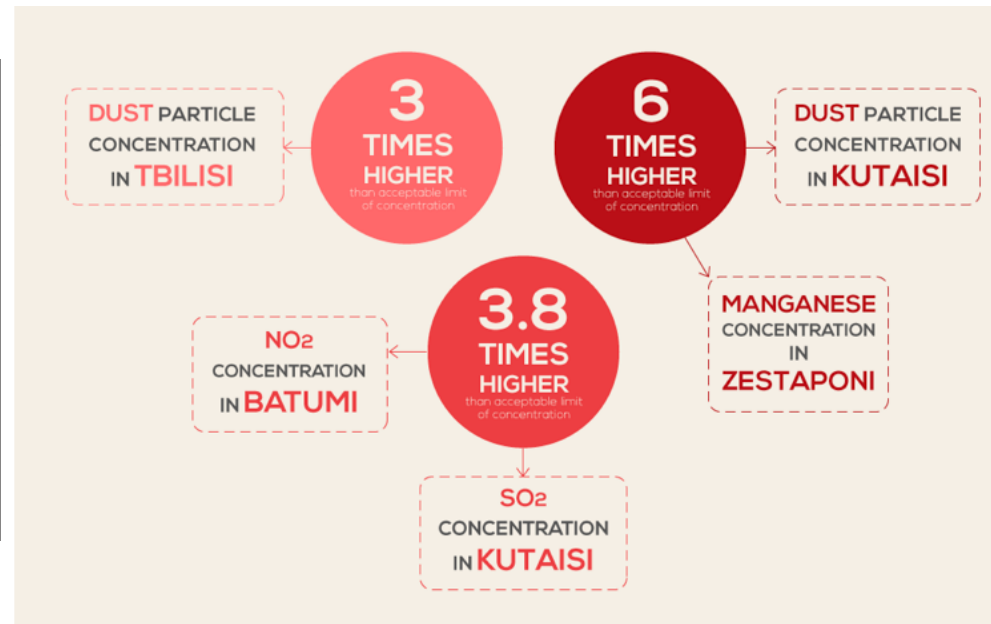
Visualized Statistics VS Data Story



- Effective title
- Concrete message
- Data to support the main message
- Providing context
- Explaining importance of data
- Creating a story flow

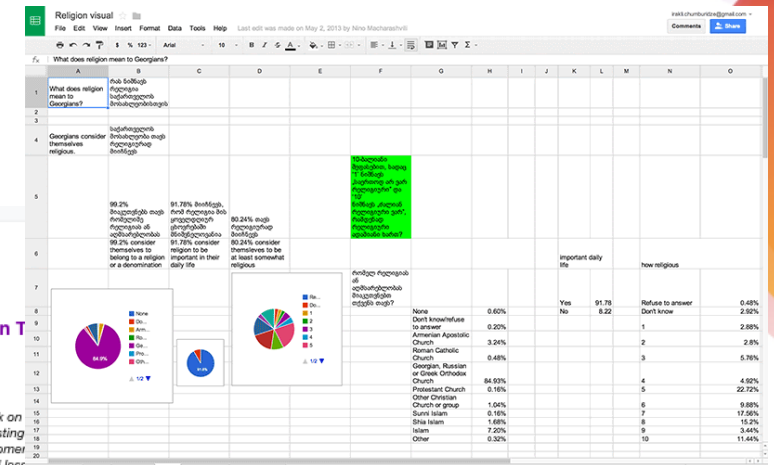


<https://jsferadi.forset.ge/en/visualizations/caution-air>



What's next?

1. Concept (feedback, revision)
2. Sketch (feedback, revision)
3. Design (feedback, revision)
4. Publishing & distribution
5. Receiving feedback & learning



“Tolerate and do not complain” – why do women in Turkmenistan endure violence?

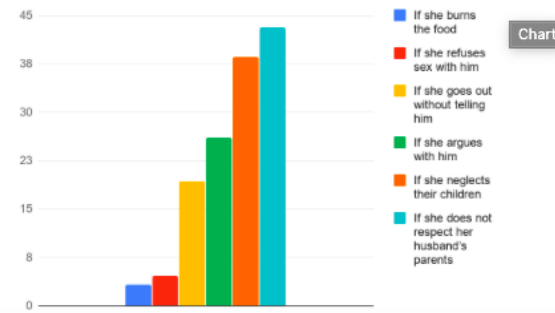
Article text:

Turkmenistan is a traditional country with a conventional patriarchal outlook on roles, where women are seen as caretakers and men as providers. The existing families perpetuate the imbalance between men and women and shape women's surface, women are expected to engage more with children and home and less women find themselves in an unfavorable position that limits them to make complaints and protect themselves. Such traits aggravate the problem of domestic violence. Despite a widespread practice of domestic violence, victims experience high community pressure thresholds to file complaints and social stigma that leads to isolation, self-blame, and internalized oppression. While left alone, the circumstances force women to believe domestic violence is normal.

INFOGRAPHIC:

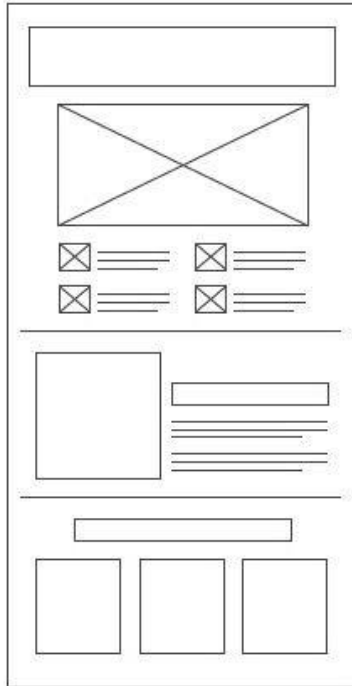
Every second woman (59%) in Turkmenistan believes wife beating is justifiable. This is almost a double increase since 2016 when it was 35%.

According to the 2019 MICS survey, women think physical violence can be justified for any of six reasons, and disrespect of husbands' parents is a more commonly justified reason for women battering:

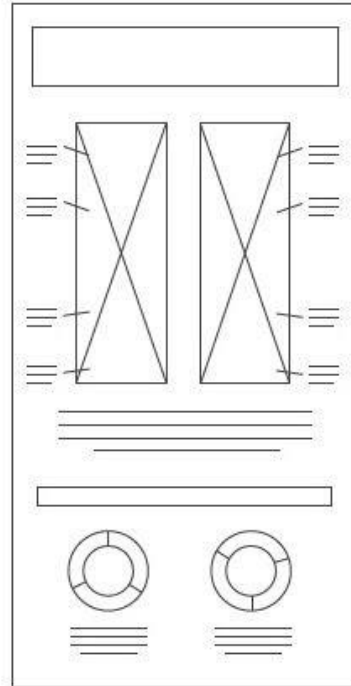


Infographic Layout Cheat Sheet

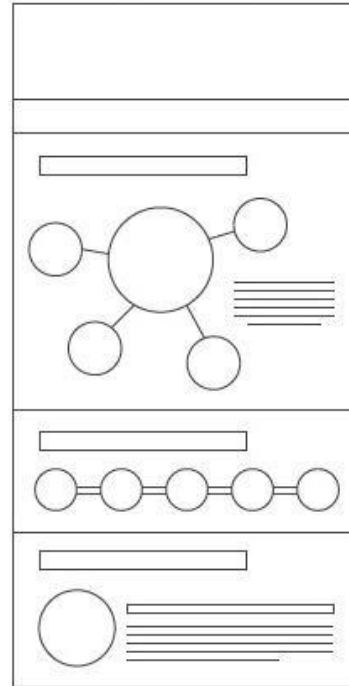
by SeeMei Chow



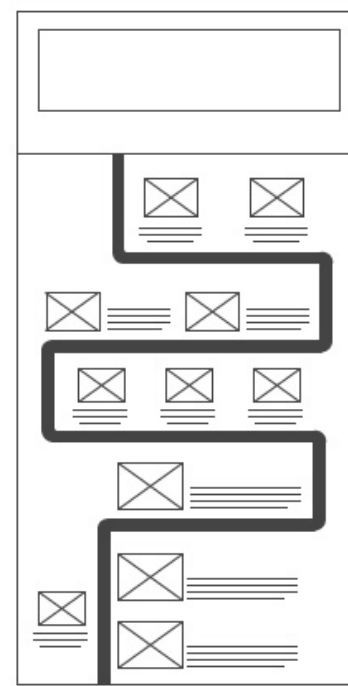
Useful Bait
Works well with most of the data
Easy to read and good usability



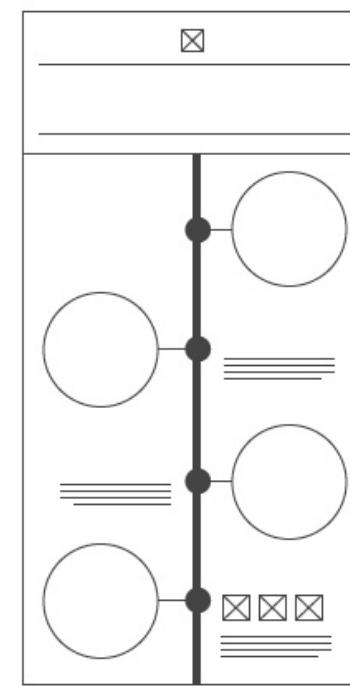
Versus/ Comparison
Works well with a lot of informations
Design(visual) is very important
Informations have to be very interesting



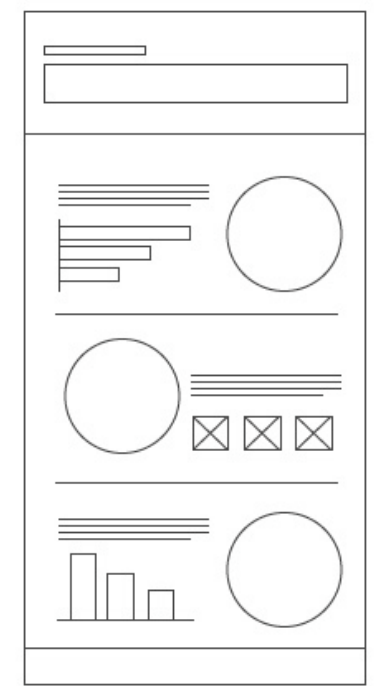
Heavy Data (numbers porn)
Works well with marketing strategy
Timeline for project
Can extend to a flowchart



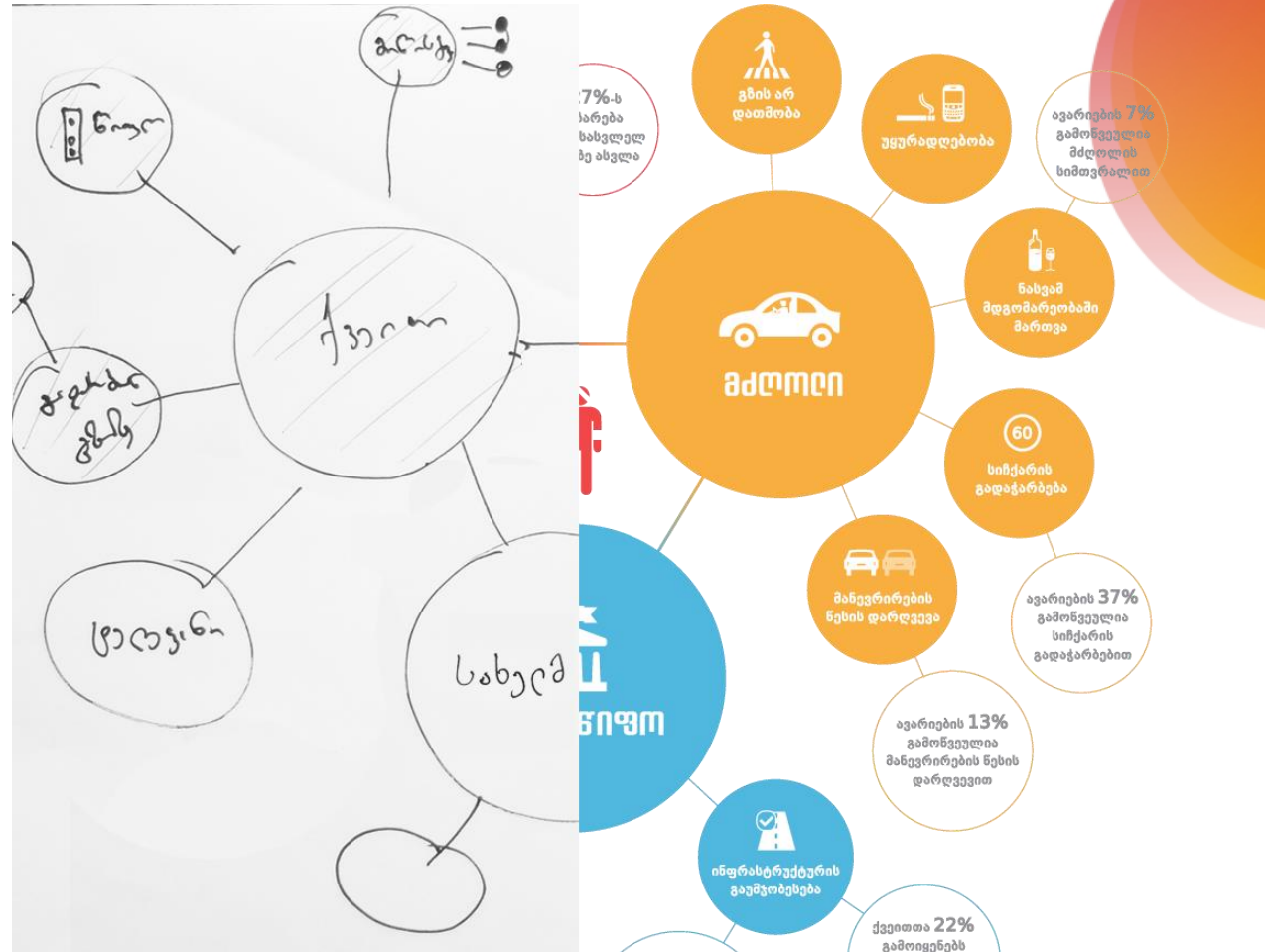
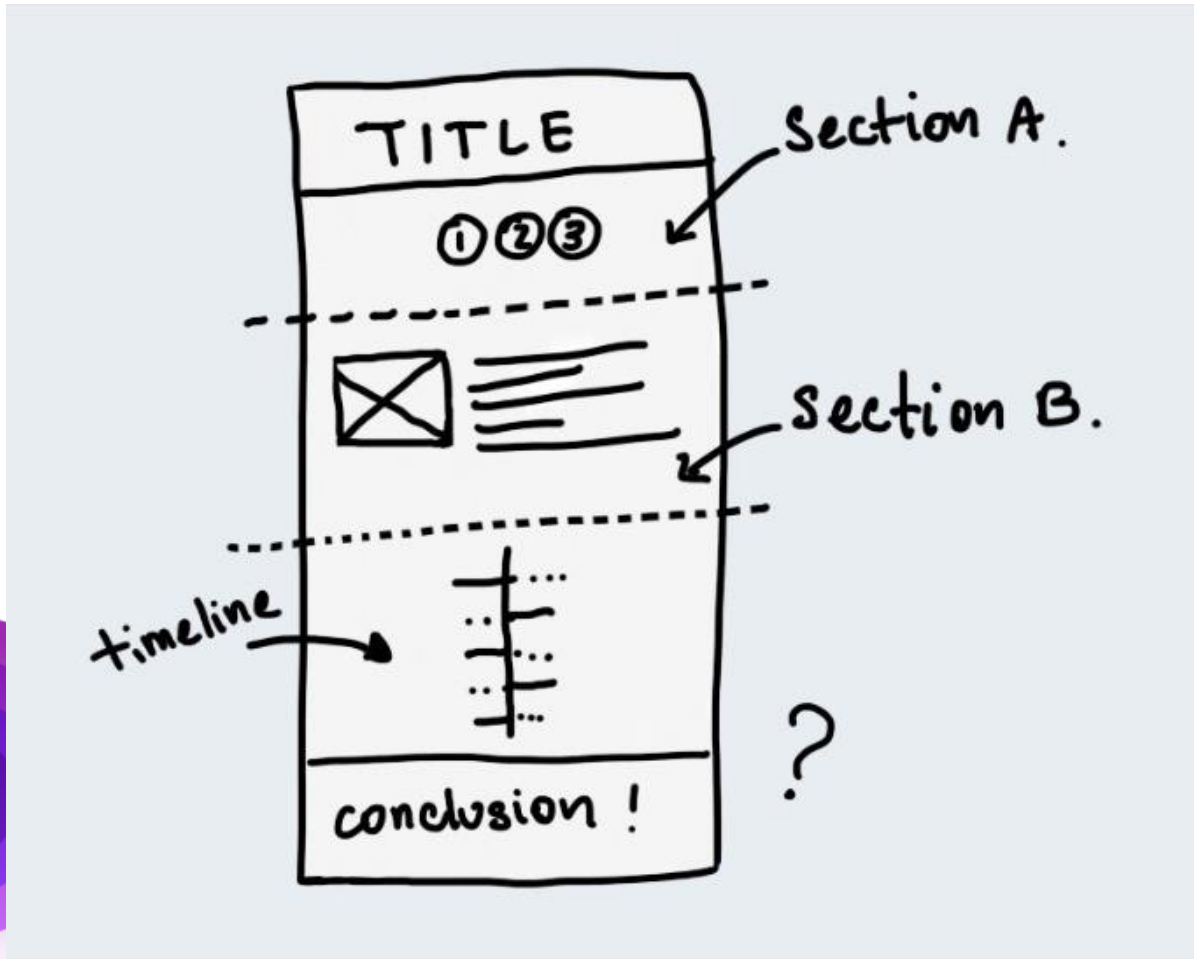
Road Map
Good for storyline/journey
Can be used as a timeline too



Timeline
Can be a comparison
Good for timeline or journey too
From simple to complex
(depends on your data)



Visualized Article
Needs strong title
Works well with heave content
Easy to read and understand



Questions?

Resources

Organizations and AI Initiatives

- Data Science For Social Good Foundation
<https://www.datascienceforsocialgood.org/>
<https://www.solveforgood.org>
- DataKind <https://www.datakind.org>
- Data Collaboratives <https://datacollaboratives.org>
- Open Data's Impact <http://odimpact.org/index.htm>
- The Humanitarian Data Exchange
<https://data.humdata.org>
<https://centre.humdata.org>
- Google AI <https://ai.google>

Communities

DataFest Tbilisi

<https://www.facebook.com/DataFestTbilisi>

Open Data and Data Science Groups

- Data Science Tbilisi
<https://www.facebook.com/groups/DataScienceTbilisi>
- Open Data Belarus
<https://www.facebook.com/groups/opendataby>
- Armenian Data Science
<https://www.facebook.com/groups/732676026823922>
- Data Science Community Azerbaijan
<https://www.facebook.com/groups/1433534440051100>
- Data Science Moldova Community
<https://www.facebook.com/groups/1820913141514829>
- E-democracy and Open Data Ukraine
<https://www.facebook.com/groups/1547226408897896>

Resources

Data sources

- World Bank: <https://data.worldbank.org>
- World Health Organization: <https://www.who.int/data>
- United Nations: <https://data.un.org>
- European Union: <https://ec.europa.eu/eurostat/>
- <https://www.asktheeu.org>
- Google: <https://datasetsearch.research.google.com>
- Open Corporates: <https://opencorporates.com>

For inspiration

1. <https://visual.ly/view>
2. <https://datajournalism.com/awards>
3. <https://jsferadi.forset.ge/en/visualizations>
4. <https://jsvisuals.forset.ge>
5. <http://dailyinfographics.eu>
6. <https://flowingdata.com>
7. <https://coolinfographics.com>

Google Advanced Search:

https://www.google.com/advanced_search

Invitation 😊



<https://datafest.ge>

Thank you for your attention!

Nino Macharashvili | ForSet
nino@forset.ge

Let's stay in touch:

<https://twitter.com/NinoMacha>

<https://www.linkedin.com/in/nino-macharashvili-b4307817/>

<https://www.facebook.com/nino.macharashvili.18>