A3 Write-up: Team DataMinders

Process & Design Decisions

I. Vision

Topic of interest: Mental Health

We started by asking ourselves what aspects of mental health were we most curious about in order to set the overall objective for A3. We identified three possible types of stories we could tell through data visualization:

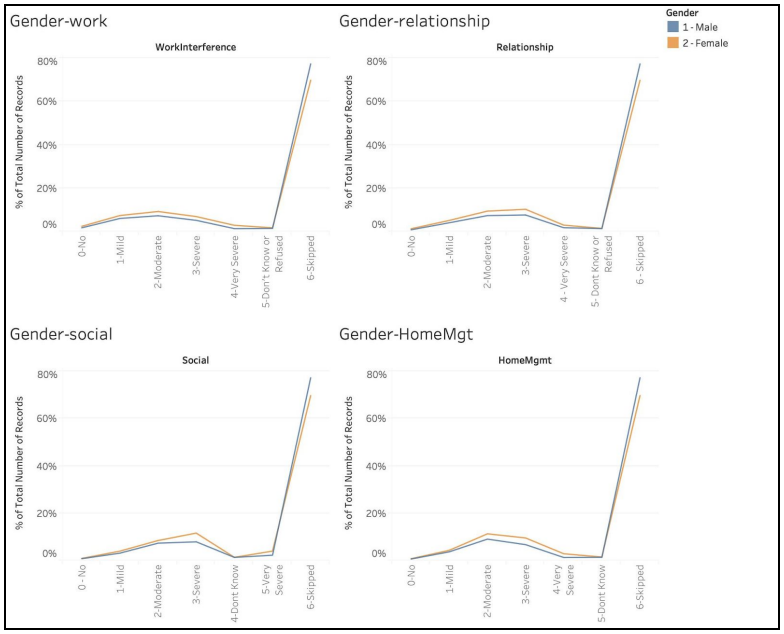
1. Analyse mental health data to identify interesting patterns and insights for behavior change, e.g. the effect of technology use on mental health
2. Analyse data to draw attention to a strategic or business problem e.g. existing gap in mental health access
3. **Use data stories to spread awareness around mental health, particularly to de-stigmatize and build empathy.**

II. Steps

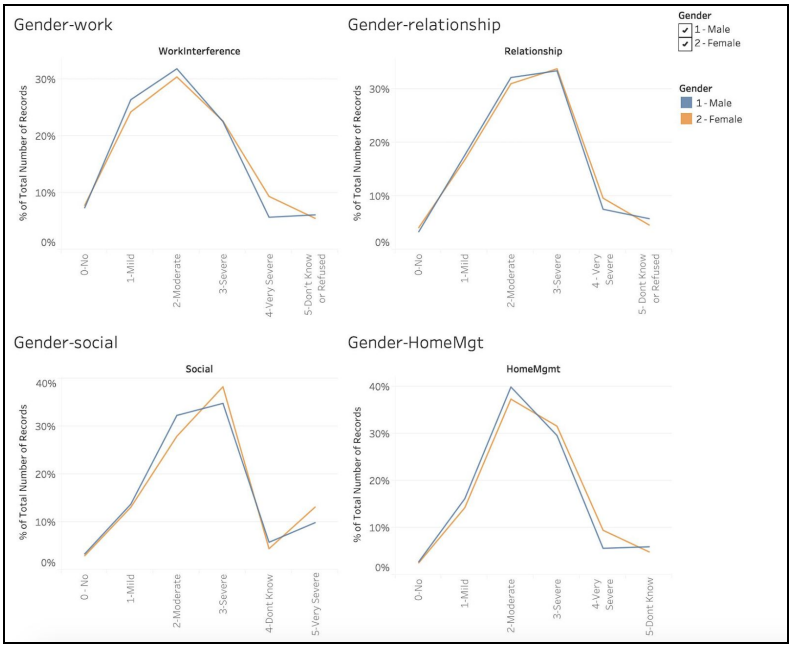
1. **Find + explore datasets:** A number of publicly available datasets were explored (Appendix A). The process involved many back and forths between the available datasets and the three directions identified as fit for the vision since we needed to make sure we can find data to represent our story. Datasets were checked for quality and analysed for variable correlations. The following is a brief list of our findings that helped us close in on our direction:
    a. We found good data around mental health prevalence around the world, country-wise, but we realized that a lot of spatial and temporal visualizations were already available. Additionally, a lot of global data available was already aggregated.
    b. For analysing mental health access gap, while we could find pre-aggregated data for mental health resources (healthcare centers and providers), it was really difficult--probably due to privacy issues--to find good data for healthcare utilization and claims data to map demand. Cost data were also very difficult to find.
    c. For better understanding people with mental illnesses, we found a couple of very interesting person-level datasets (populated through national health surveys) that sampled US households to track healthcare status and habits. In addition to the datasets in this area being very rich, we also realized that these stories have rarely been told or visualized. To us, visualizing the lives of people who experience depression was the most meaningful. ***We therefore decided to select the third vision as the project vision and utilized one of these datasets.***
    (https://www.datafiles.samhsa.gov/study-dataset/national-survey-drug-use-and-health-2016-nsduh-2016-ds0001-nid17185)

2. **Select/filter relevant subset:** Our team spent a long time exploring what the average interference level for those with depression looks like based on different factors. There were clear correlations in certain categories. For example, women as a whole had a higher average depression interference score compared to men. However, we realized that the nuances in individual stories are lost if the stories are grouped together and aggregated (shown through charts below). Therefore eventually, instead of looking at aggregate data, we sought to represent the very different narratives of people who experience depression on an individual level. The dataset had 5 effects of depression that made sense to us to further explore - interference with work, social life, home management, and relationships, as well as number of days in the last year that were lost due to depression.

   Dashboard1: The following group of charts show prevalence of depression interference (severity-wise) for women and men in the entire survey group. We see that women overall have reported higher interference in their life due to depression.



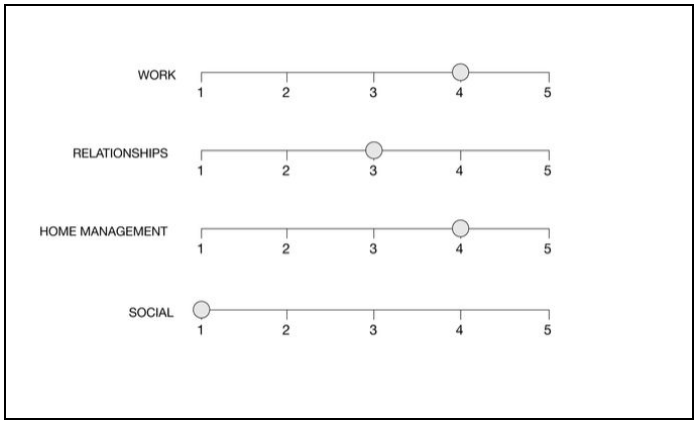
Dashboard 2: Note, how the severity differences vanish when only people who were diagnosed with depression are considered (and not the entire universe).



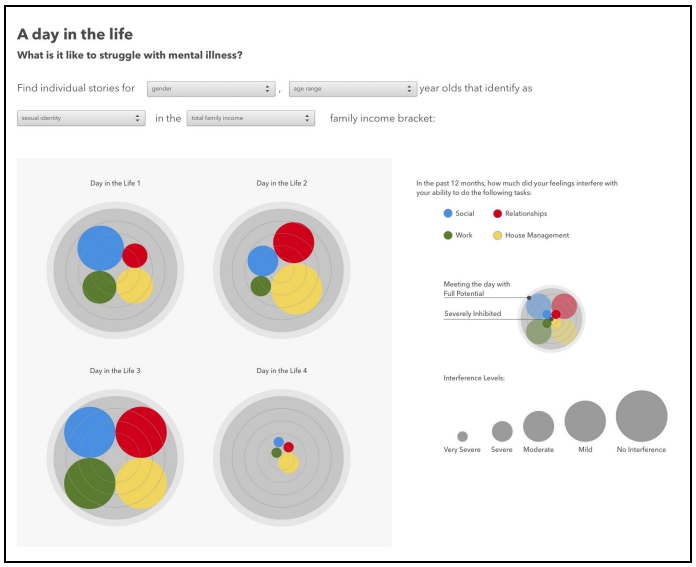
Choosing demographic information of interest: We developed multiple charts that analysed depression correlation with different variables - age, gender, employment status, marital status, sexual orientation, income, region, employer etc. We found strong correlations with most variables. In the end, we decided to select age, gender, income, and sexual orientation as our filters, because we thought those are the variables with the most variation and that people most relate to.

3. **Mockups:** Having decided the story we wanted to tell and variables we wanted to show, we started mocking up the visualization in different ways.

   Sliders: Initially, one way we thought we could visualize depression's impact on one's life was through four moving sliders from 0 (meaning no interference) to 10 (meaning severe interference). However, it would give the impression that the sliders can be adjusted by hand even though the graph could only be changed through external filtering.



Bubble Chart: We also thought of bubble charts depicting the same concept. Each bubble would represent a part of life with the size of the bubble telling us the severity of difficulty in performing tasks for that aspect.



Polar Chart: The concept that resonated with the team most revolved around “completeness” in one’s life. A circle that consumes the entire plot represents having no interference from depression in life activities--a full domain of ability. As the effects of depression increase in each aspect, however, the area of the circle shrinks and deforms, showing the shrunken domain of one’s ability to perform tasks in each category.

4. **Final Design/output:** Detailed below.

III. Contribution/Output

1. Final Question

What does depression look like? What does it mean to be depressed? We wanted to convey the idea that depression as an illness is not well understood. People often confuse sadness, the emotion, with depression. As a result, if someone is suffering from depression, others quite often write it off as a temporary emotion not realizing that depression is much more than that. It can be extremely debilitating and serious. Often, the lack of information results in a lack of support and empathy for someone with the very real effects of depression.

2. Final Concept

We decided to answer the above question by analysing and visualizing the magnitude of interference that depression causes in an individual’s life, life as defined by four important parts - one’s work, social life, relationships, and home life, as well as showing how many days the person could not function in those 4 areas due to depression. In

order to do so, given a combination of factors one can enter (age, gender, sexual orientation, and family income), four people matching that description are selected at random to provide a sample of the different experiences and challenges they face.

3. Visualisation

Radial Polar Chart We used a radial polar chart for our interactive visualization since we could depict the four parts of our analysis (work interference, social interference, relationships, home management) on the four axis and visualize a resulting shape from the four values. Furthermore, we decided to inverse the scale, so that a response of 0 would mean maximum severity whereas a response of 10 would mean 0 interference. We did this because, we wanted to communicate that with increase in severity, one's life gets more and more limited and therefore the circle that was once full, distorts and shrinks.

In terms of effectiveness, we were inspired by "The Creative Pace of the 20th Century's Greatest Authors, Visualized" from class in a sense that we wanted the viewer to take a moment to ponder and understand the legend/encodings but at the same time communicate the point of the visualization quickly once the system was understood. We still designed it so as not to overwhelm the user with too much information to process.

a. Visual Channels We used position on a common scale to visually encode the magnitude of ability to perform tasks in specific categories. Since we used a polar chart, connecting the four data-points resulted in a shape that shrank as the ability to perform in a specific category was reduced. The resulting shape varied in luminance and saturation based on the number of days that the person was not able to work/carry out daily activities due to depression. The number of days were binned into 6 categories so the variance in luminance and saturation can be more easily recognized.
b. Graph Labels The labels provided the specifics of each person in the group's life, in addition to the exact number of days the person was unable to carry out work due to depression. Since the values for days lost were binned for a more powerful recognition of the magnitude, we decided that showing the actual number would help complete the story.
c. Legend The legend clarifies the two types of data each plot represents: success in participating in four different areas of life, and days lost, both with respect to depression. For days lost, the legend defines a strategic gradient from light blue, like a sunny day, to dark grey, reminiscent of a stormy sky. We realized that over 70% of all people affected by depression have at least missed one day of work due to their depression. By showing the legend as the brightest possible value indicating 0 days missed, we can emphasize the stark contrast between the mostly desaturated colors that constitute most of the stories.

Commentary about the development process:

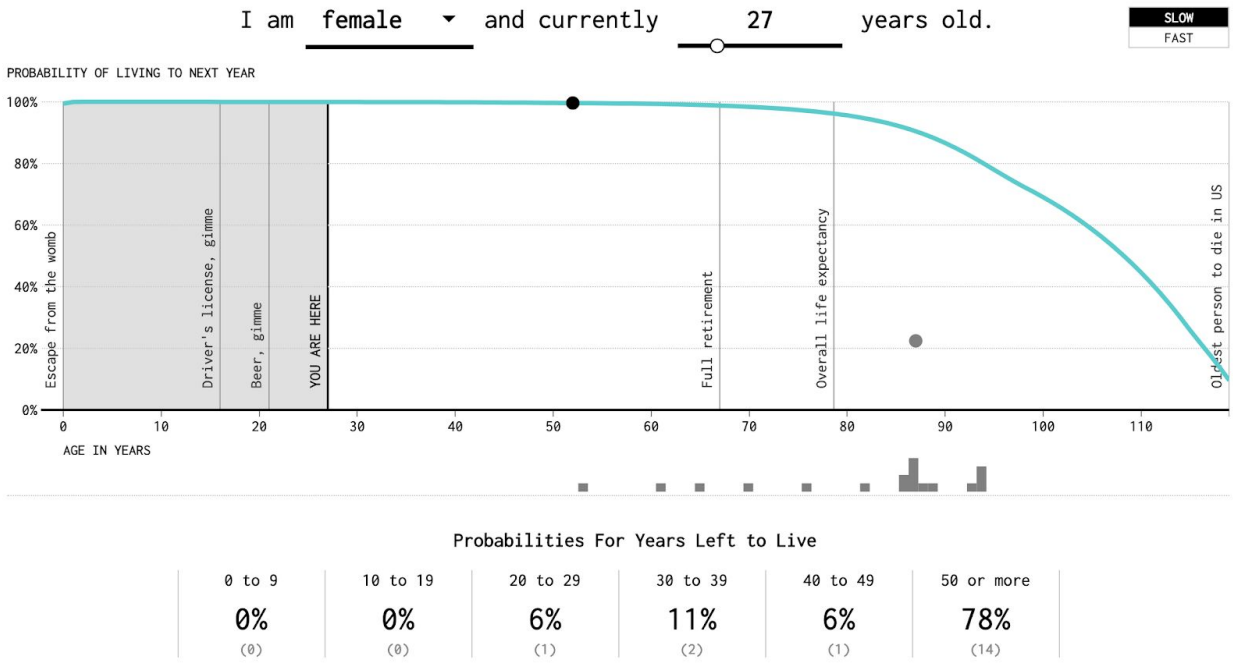
Task	Sub-Task	Person(s) Responsible	Time Taken (person-hours)	Notes
Initial Brainstorming	Visual Inspirations	Andrea, Anubhav, Sid, Nava	4	
Finding Relevant Dataset and Subset / initial exploration	Data wrangling	Andrea, Anubhav, Sid, Nava	12	Finding good, clean, recent, and reliable dataset was more challenging than expected
	Data Exploration	Andrea, Anubhav	30	Since we were exploring a bunch of datasets, this step took us a lot of time. Also, a lot of datasets were not clean and had to be processed in order to be explored.
	Preparing Final Dataset	Andrea, Anubhav	8	Did this in R which took some time to learn
Static Visualization	Inspiration +Ideation + Sketches + Mockups	Nava, Andrea	20	This step took longer than expected since as new design concepts were explored, we realized we had to look for a different dataset than what we had finalized.
Interactive Visualization	HTML + D3 framework	Sid	40	Learning D3.js took up a significant amount of time in making this visualization. This step took as much time as expected.
	Scatter Polar Chart + legend	Nava	15	
	UI/UX/CSS	Andrea	10	
	Debugging + Final Touches	Andrea, Anubhav, Sid, Nava	20	Debugging took a long time. For example it took us some time to figure out how to address as many edge cases as possible for filtering. Other debugging included customizing existing libraries.
Write-up		Anubhav, Nava	8	
Total			167	Overall the surprising step that what took the longest was going between finding datasets, sorting and cleaning them up, exploring the set, understanding the possibilities, ideation and mockup based on those possibilities, then going

				back and finding the appropriate dataset.

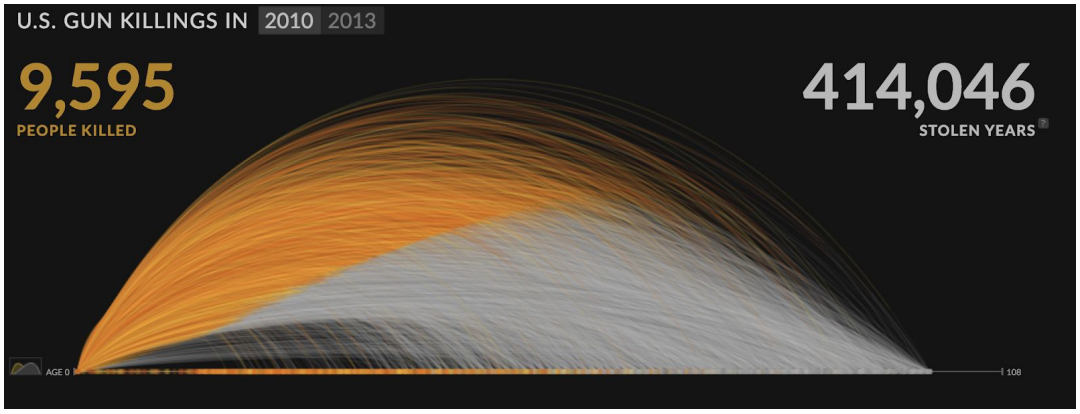
Appendix 1: Datasets Explored

1. https://www.cdc.gov/brfss/index.html: Aggregated data about mentally unhealthy days, anxiety and depressive disorders, mental illness and stigma, and psychological distress.
2. https://data.world/health/health-related-quality-of-life: Similar to above, aggregated location-wise prevalence data on depression
3. https://ourworldindata.org/mental-health: Global aggregated spatial and temporal data on mental health prevalence
4. http://ghdx.healthdata.org/gbd-results-tool: Interesting aggregated dataset on mental health prevalence grouped by age, race, gender, country etc.
5. https://findtreatment.samhsa.gov/: Data on mental health access - geographic distribution of mental health centers and providers
6. https://www.cdc.gov/nchs/nhis/nhis_2017_data_release.htm - person level healthcare behavior data and status. Has information about depression and other lifestyle characteristics including days lost, internet, meditation etc. However, depression variables can't used to map out prevalence
7. https://www.datafiles.samhsa.gov/study-dataset/national-survey-drug-use-and-health-2016-nsduh-2016-ds0001-nid17185 - our final dataset that has person-level information about various lifestyle and health behaviors including depression and its impact. The National Survey on Drug Use and Health (NSDUH) series, formerly titled National Household Survey on Drug Abuse, is a major source of statistical information on the use of illicit drugs, alcohol, and tobacco and on mental health issues among members of the U.S.

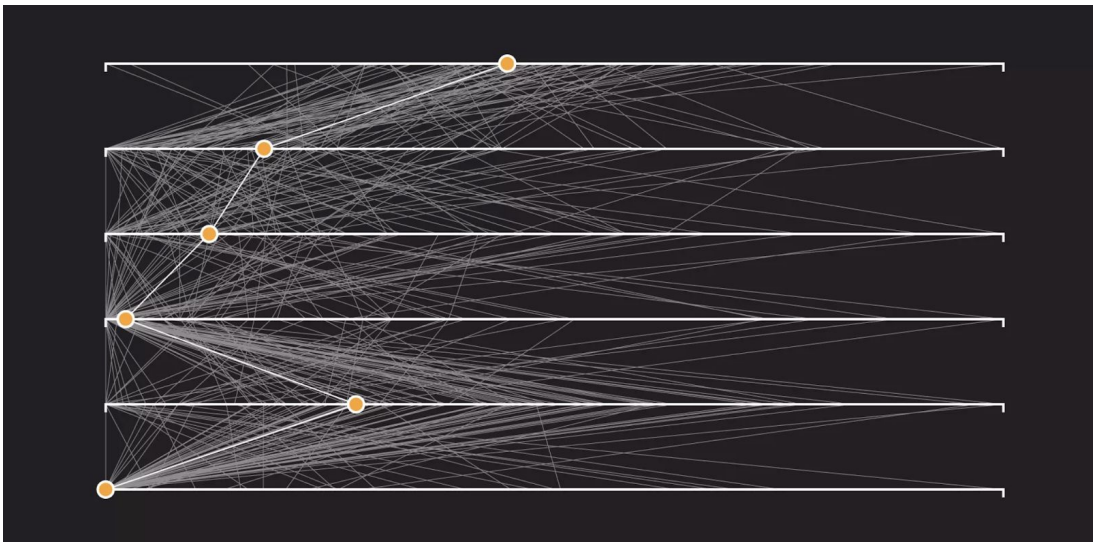
Appendix 2: Visualization Inspirations: How can data tell stories?



1. https://flowingdata.com/2015/09/23/years-you-have-left-to-live-probably/

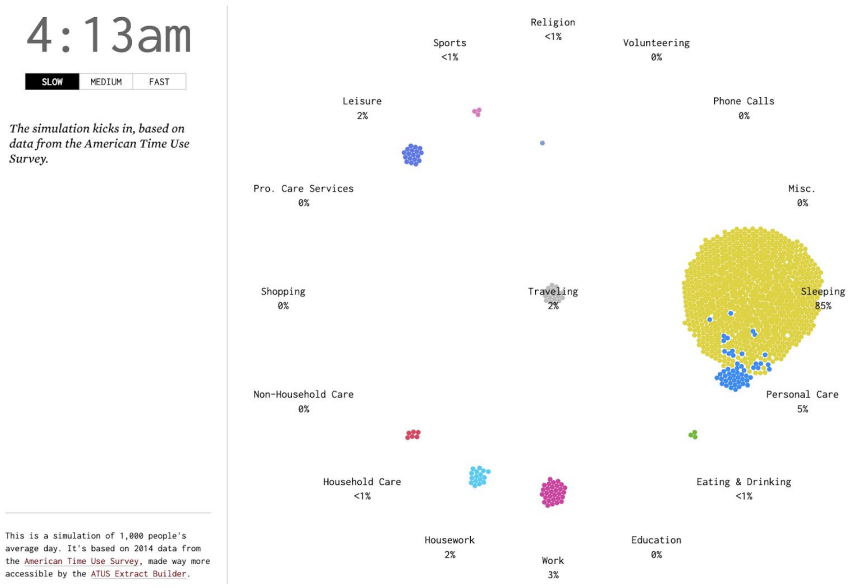


2. https://guns.periscopic.com/?year=2013

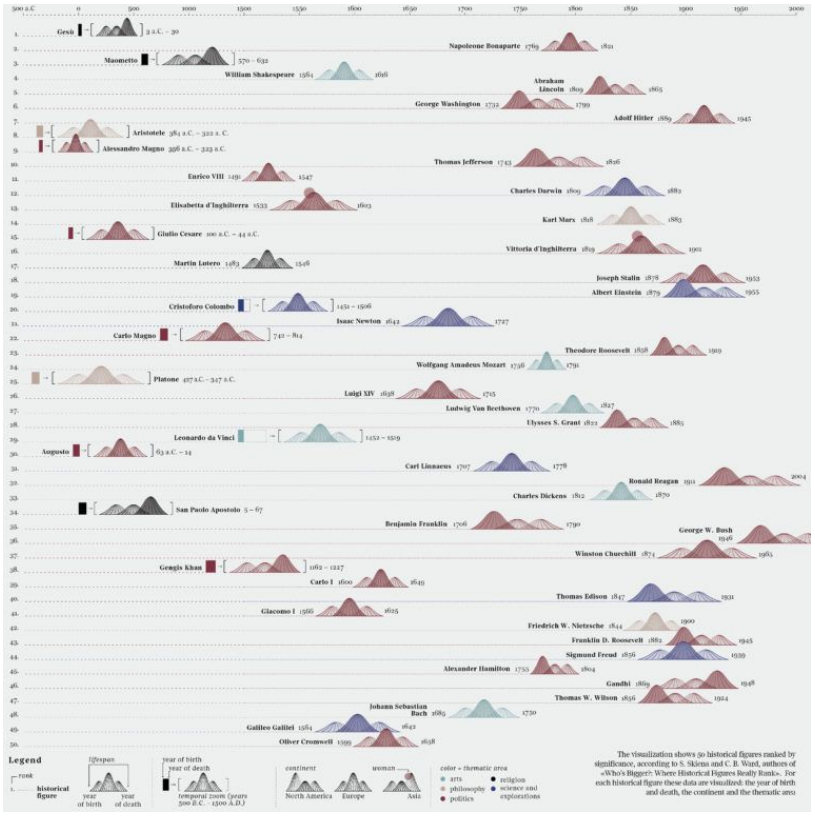


3. https://flowingdata.com/2016/12/06/how-people-like-you-spend-their-time/

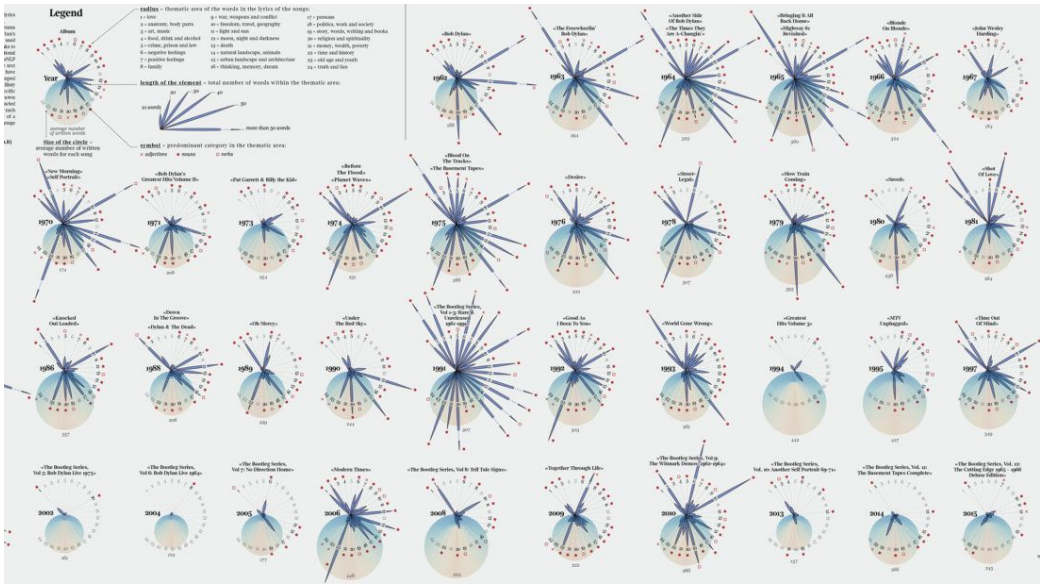
4. https://informationisbeautiful.net/visualizations/who-old-are-you/



5. https://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/



6. https://europeandesign.org/submissions/data-visualizations-for-visual-data-corriere-della-sera-la-lettu


Legend
Album
Year
average number of written words
Slot of the circle – average number of written words for each song
radius – thematic area of the words in the lyrics of the songs:
1 – love
2 – anatomy, body parts
3 – art, music
4 – food, drink and alcohol
5 – crime, prison and law
6 – negative feelings
7 – positive feelings
8 – family
9 – war, weapons and conflict
10 – freedom, travel, geography
11 – light and sun
12 – moon, night and darkness
13 – death
14 – natural landscape, animals
15 – urban landscape and architecture
16 – thinking, memory, dream
17 – persons
18 – politics, work and society
19 – store, words, writing and books
20 – religion and spirituality
21 – money, wealth, poverty
22 – time and history
23 – old age and youth
24 – truth and lies
length of the element – total number of words within the thematic area:
10 words
30
39
49
59
more than 59 words
symbol – predominant category in the thematic area:
○ adjectives
● nouns
□ verbs
«Bob Dylan»
«The Freewheelin' Bob Dylan»
«Another Side Of Bob Dylan»
«The Times They Are A-Changin'»
«Bringing It All Back Home»
«Highway 61 Revisited»
«Blonde On Blonde»
«John Wesley Harding»
«New Morning»
«Self Portrait»
«Bob Dylan's Greatest Hits Volume II»
«Pat Garrett & Billy the Kid»
«Before The Flood»
«Planet Waves»
«Blood On The Tracks»
«The Basement Tapes»
«Desire»
«Street-Legal»
«Slow Train Coming»
«Saved»
«Shot Of Love»
«Knocked Out Loaded»
«Down In The Groove»
«Dylan & The Dead»
«Oh Mercy»
«Under The Red Sky»
«The Bootleg Series, Vol 1-3: Rare & Unreleased 1961-1991»
«Good As I Been To You»
«World Gone Wrong»
«Greatest Hits Volume 3»
«MTV Unplugged»
«Time Out Of Mind»
«The Bootleg Series, Vol 5: Bob Dylan Live 1975»
«The Bootleg Series, Vol 6: Bob Dylan Live 1964»
«The Bootleg Series, Vol 7: No Direction Home»
«Modern Times»
«The Bootleg Series, Vol 8: Tell Tale Signs»
«Together Through Life»
«The Bootleg Series, Vol 9: The Witmark Demos 1962-1964»
«The Bootleg Series, Vol. 10: Another Self Portrait (69-71)»
«The Bootleg Series, Vol. 11: The Basement Tapes Complete»
«The Bootleg Series, Vol. 12: The Cutting Edge 1965 - 1966 Deluxe Edition»