ETC5543 - Business analytics creative activity - S1 2023

Internship Report

Nishtha Arora (32296622), Monash University, Victoria, Australia

2023-06-15

Abstract

For decades, computer based reporting has been an integral part of journalism, that uses public records, databases, private and public data sources to investigate patterns, trends or even anomalies in the data collected. The integration of data analysis in the reporting industry brings challenges with it, i.e., data manipulation, wrangling, access to platforms supporting visualization reorganization, etc.(Halevy & McGregor, 2012). The aim of the project is to support the journalism team of authors and editors with compelling visualization to support their claims and research, or creation of an analysis via a visual related to the topic selected. The first package, **'Decriminalizing Suicide'** focuses on various aspects covered by the authors, one of them is 'India's Mental Health Act 2017' and opposite results are observed due to different data sources, an increase and decrease after 2017, which is discussed in the report. The second package, **'Policing he Police'** takes a generic angle, covering topics of shootings around the world, the changing trust in police and much more. The project uses methods of data wrangling, exploration, pdf scraping, spatial analysis and basics of functions of tidyverse, in R language and uses 'themes360info' package for the theme. The report is divided into four main sections, i.e., Introduction, Aim, Methodology and Results/Learnings. The workflow for both the selected packages, starts with initial analysis for the topic, the shortlisted/selected visualizations with 360info theme added, reasons for selection or rejection of a particular visualization and the challenges faced during the process.

Contents

About

360info is a not-for-profit open access agency that provides global information regarding world's issues and provides solutions for the same. This content is forwarded to re-publishers without charge, under Creative Commons.

The content published is based on research, and each week a special report is published, focusing on a global problem, which consists of 5-10 articles covering different aspects in the problem. These articles are contributions from academics across various fields of study, depending on the article.

Each report is supported by visuals, can be images, graphics or interactives. Any story telling can be made better with a data-driven analysis along with it and hence, this internship has given me a chance to work in the data and digital story telling team, produce data visualizations, collaborating with the authors and editors.

All the published work is reproducible for media partners and is published under Creative Commons licences, which is good for art, educational and entertainment works. 360info uses Creative Commons attribution 4.0 because it allows the user's rights under the licence to be reinstated, if the user comes in compliance within 30 days of discovering that they were in violation of the rights.

Background and motivation

Suicide is a worldwide public health problem. There have been over 700,000 deaths from suicide worldwide in 2019. Overtime, there have been a no. of theories if decriminalizing suicide is a boon or a bane? Will it increase the suicide rate or decrease it? It may decrease the overall rates because then people will start talking about it openly, which will improve mental health and therefore, less suicides, or it may increase the attempt to suicide rate. According to WHO, there are still 20 countries that have criminalized suicide (World Health Organization: WHO, 2021).

The British common law stated that one has no right to take his/her life as it belong to the state and this affected many former British colonies like Kenya, who still criminalize suicide, even after the colonization ended. The Christian Commandment of 'Thou shall not kill' signifies that one should not kill himself/herself as will. And Suicide is a sin under the Sharia Law, under the Islamic Tradition (Ochuku et al., 2022).

With the advancement of science in the 19th and 20th century, it was discovered that suicidal tendencies are caused by biological factors as well and hence, continents like Europe and North America revoked the laws regarding criminalize suicide. Further, as the years went by, and awareness increased, lot of policies came into action, like Convention on the Rights of Persons with Disabilities and World Health Organization Mental Health Action Plan 2020–2030 prompted various countries to decriminalize suicide

Suicides are are result of no. of causes, ranging from abuse victims, loss, loneliness, use of intoxicants to financial issues. All these issues result in mental breakdown and it is safe to say that all potential suicide victims go through a mental health issue, it might not be true vice versa. These mental health issues come with stress, anxiety or depression and often times are linked to suicidal feelings or behaviour and might not be the only cause of suicide. The relationship between mental health and suicides is complex.

Theories like 'criminalizing suicide prevents people from reaching out for help which results in an increase in suicide rate' or 'criminalizing suicide would decrease the attempts made to suicide and hence, lower the suicide rate', are up for debates.

The concept of **Policing the Police** has been emerging recently because of historic law enforcement officers not caring about and allowing misconduct by the police, due to less resources and external power. There was no check kept on the Police, which lead to a no. of reforms and protests by the police. An example of this is the death of George Floyd, an African-American man murdered by a police officer in Minneapolis, Minnesota, over Floyd being a suspect for using a counterfeit twenty-dollar bill.

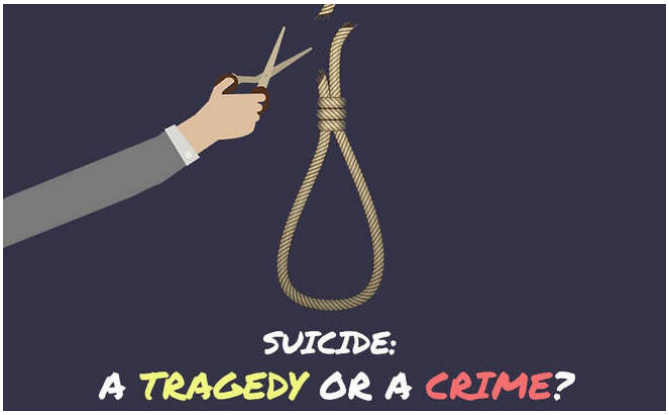
Police is responsible for our safety and we rely on them for protection. But it is not everytime, that thye can be trusted nowadays. 'Police accountability' is up for debates around the world. With about 1440 cases recorded against police within England and Wales during 2019-2020, 3.4% complains against law enforecemnt officers involving racism and discrimination in Australia and mass shooting in the US, concern regarding police integrity check and punishments are being discussed around the world.

There are certain plans and implementations that are being enforced for the process like independent investigations per officer, body cameras that provide proof of the misconduct, public surveys and strickter punishments. It is being stated that those who are supposed to protect us, must also be overseed by a body to keep actions of law in check.

There are debates about increasing reinforcements on the police, to improve trust and accountability which some may argue that increasing oversight on the police, may bring down the police morale and may affect police efficiency.

Here, is is very important to discover, what all are the instances where the police is involved and needs to be checked. Also, it is important to discover the factors that might lead to police misbehaving, whether it is in the corruption department, killings, etc.

Objectives and Significance



For every one completed suicide, 20 more attempts are made. Identification of potential suicide victims via these attempts can result in help-seeking and prevention of suicide but criminalizing it hinders the help-seeking and also results in inaccurate tracking of suicides.

For the instances and issues addressed above regarding police misconduct, the significant of this issue is vital for the growth and well being of the world. It is important to find out the relation between countries status and police mishapps, i.e. how the financial or political history may or may-not affect the occurrences in that area. Are there any patters seen over the years for a particular area in the world? Can religion be a factor in this? There are so many questions that can be answered with help of data, for a better result.

These results are important to be tackled with. On the basis of this there can be better reforms, bills and laws passed that could assure transparency, police-in-check laws. Areas of improvement can be targetted and the use of force can be monitored around the world accordingly.

Hence, the objective of this project are:

- To perform need analysis for the package 'Decriminalizing Suicide', i.e., work with the authors to make their articles stronger with statistical proof.
- To identify factors affecting and aspects related to 'Policing the Police' and work on a generic visualization, giving an idea about how things have changed overtime.
- To discover differences amongst different data sources and research to select the aligning/trusted source.
- Also, to tackle the data gaps and anomalies.

There are data gaps in data round the world for a particular year, years or season which could be a result of no. of factors like, a change in the government, a sudden technology advancement, a low economy country, major events, etc.

Methodology

Each package, i.e. 'Decriminalizing Suicide' and 'Policing the Police' would have a special report, which would contain about 8-9 articles covering different topic aligning to the package.

This project allowed us to stick to 'Static Plots' and not interactions, as for interactives to be published, plotly isn't the best tool, and javascript is preferred, which we would have had to get comfortable with, but due to time constraints and prioritizing the aim of the project, 'Static' worked for the best.

- **Step 1:** Creating initial visualizations, aligning with specific draft articles.
- **Step 2:** More relevant plots were made and shortlisted where 360themes was added.
- **Step 3:** Plot with corrected flaws was made for one of the short-listed plots. Save this plot in .png.
- **Step 4:** Create a renv.lock file using capsule package and save it to Github, for example:

```
```{r capsule, echo=TRUE}
install.packages("capsule",
  repos = c(mm = "https://milesmcbain.r-universe.dev", getOption("repos")))
capsule::create("analysis1.rmd", "analysis2.rmd")
```
```

Process carried out by Data and Story-Telling Lead, after finalizing the visualization

- **Step 5:** To transfer the image to the code of the publication directly, .png file (saved) would be considered.
- **Step 6:** To reproduce the code, the renv.lock file (created above) is used by using the following code (installing renv package and renv::restore to install the same R packages used by interns in their project)

```
```{r renv, echo=TRUE, include=FALSE}
install.packages("renv")
renv::restore()
# it will probably ask if they want to activate the project first. say 'y' to this
```
```

- **Step 7:** knit the project as usual.

Note: Every text in the report, that's italic, is a hyperlink

Data, Results and Discussion

Decrimanalizing Suicide

Initial Visualizations

Visualization 1: A generic visualization for the package Decimanlizing suicide- Crimanlizing suicide only makes it worse.

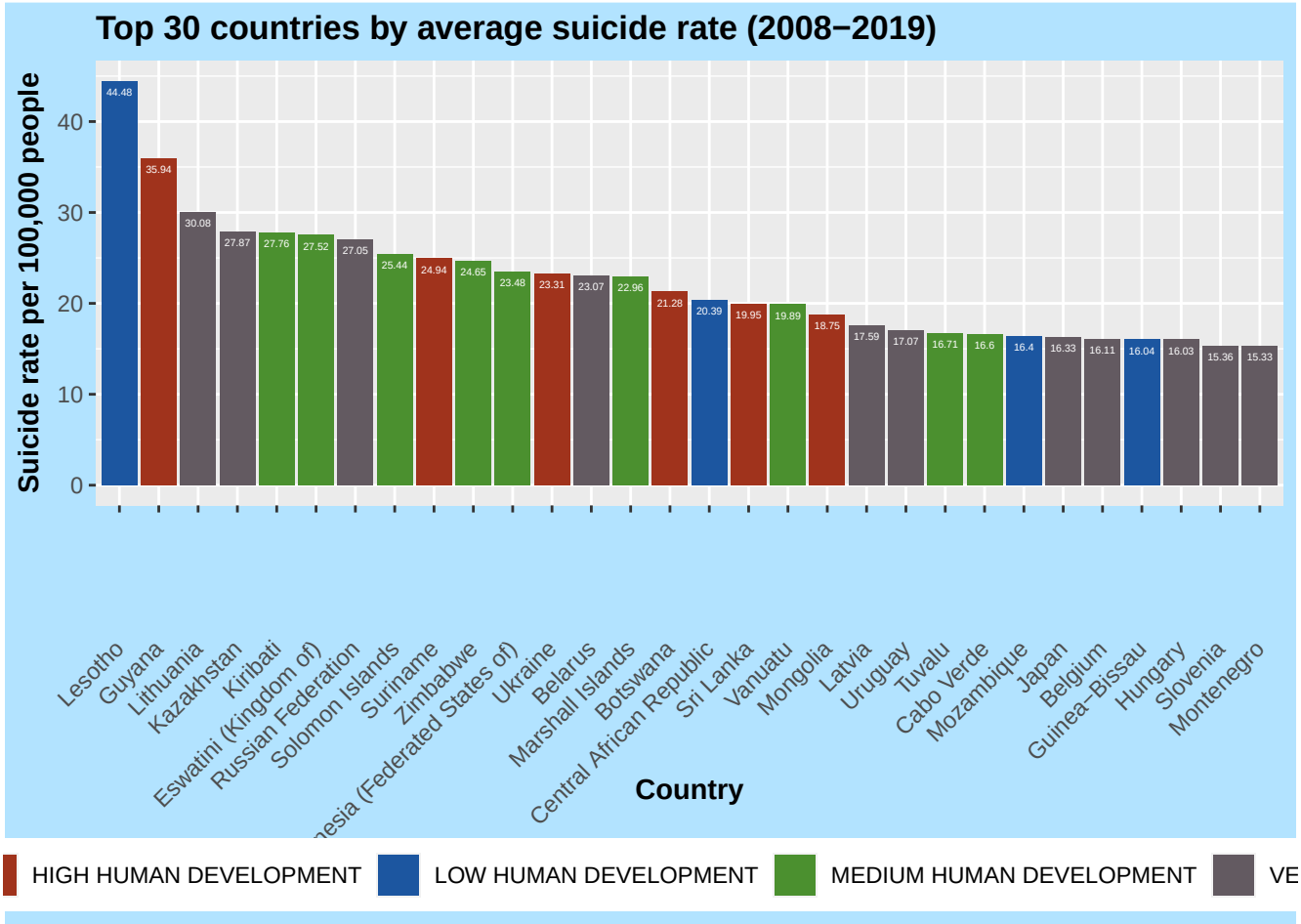
Data source:

- World wide suicide rates: Our World in Data
- Human Development Index: UNDP

HDI (Human Development Index) is a statistic composite of life expectancy, mean years of schooling, expected years of schooling and per capita income. These indicators are used to classify countries into four tiers of human development.

Table 1: Range of HDI ranks for Human Developent 2020

HDI_Value_2019	Development_Index
0.8-1	Very High HDI
0.7-0.8	High HDI
0.55-0.7	Medium HDI
0-0.55	Low HDI



Experiment for further visualizations.

For further visualizations, a more legit data source was recommended, and hence a global suicide rates data was extracted from World Health Organization which contains global data of suicide rates from 2000 to 2019..

Comparison of data from OWID and WHO by selecting a random country, say Australia.

It is observed that the WHO data values are higher than OWID values for the years 2014 and after, and lower for the years 2010-2014.

Visualization 2: This visualization observes data gaps and reduncies in the dataset and was to be paired up with What a suicide database registery should look like

Data source: Global Suicide Rates WHO

Here, the objective of the visualization is to confirm significant errors in any data and why any data source cannot be fully trusted. This is done by observing outliers in the data set. *Stephen Hawkins described Outliers as a point that deviates so much from the other observations that it arises a suspicion about a different mechanism being used for its generation*(G, 1987).

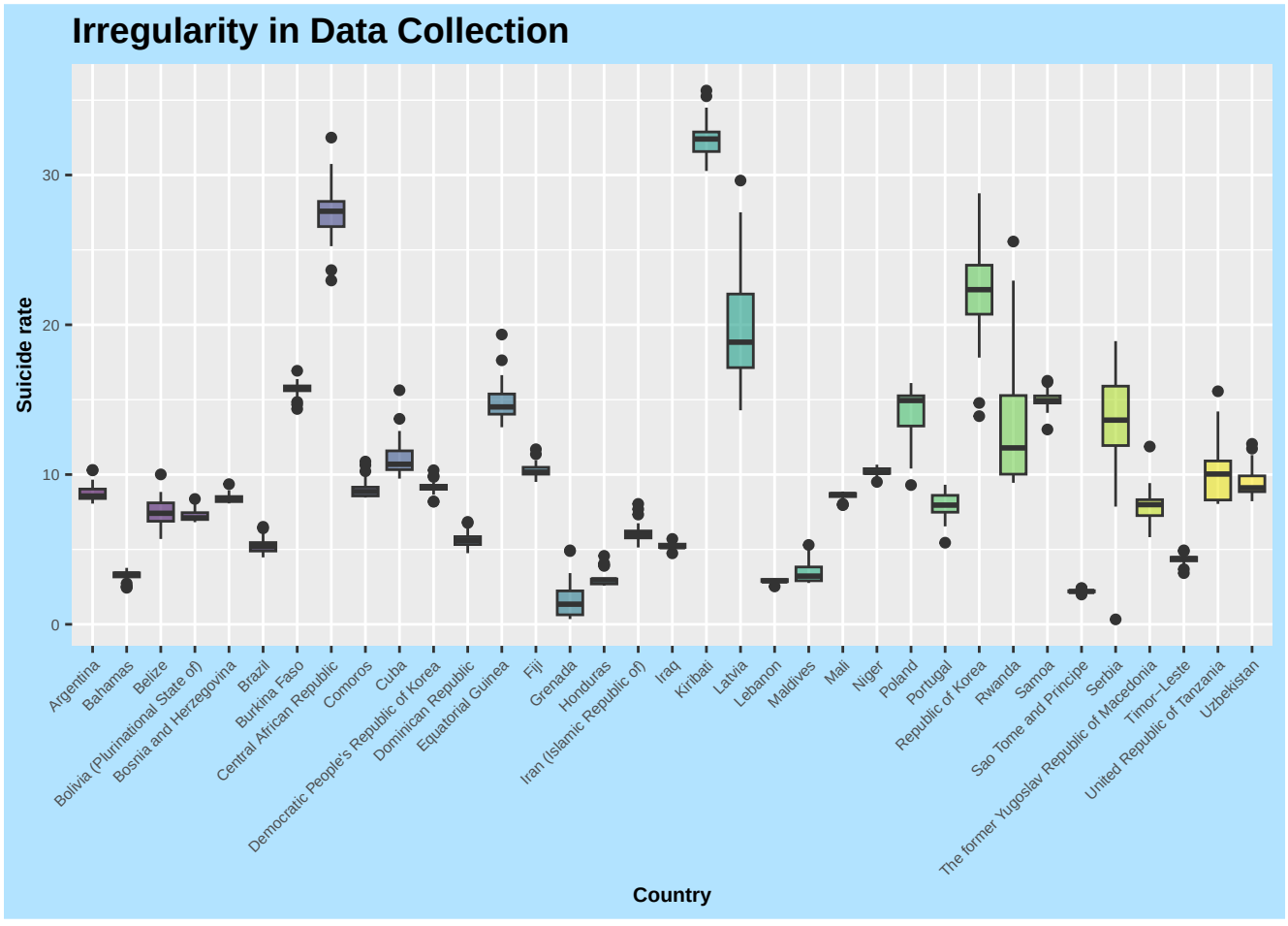
These data points vary differently and could be due to no. of reasons, for example, variability in measurement, hampering of data, misreporting, under reporting, duplication, sampling errors, unusual events, human errors of recording incorrect data or miskeyed upon data entry, etc.

Table 2: Comparison of Data Sources- Australia's Suicide rate

Year	suicide_rate_OWID	suicide_rate_WHO
2019	10.39	11.25
2018	10.44	11.26
2017	10.50	11.79
2016	10.90	10.92
2015	11.35	11.81
2014	11.11	11.31
2013	10.78	10.24
2012	10.74	10.48
2011	10.81	10.06
2010	10.90	10.41

Outliers are highly underestimated! A small proportion of outliers can affect a simple analysis, giving rise to inflated error rates and distortions in statistical estimates and removal of these can help improve the accuracy significantly(Osborne & Overbay, 2004).

Here, initially the complete data set was observed for observing outliers, but due to it being a large data set, text overlapping and squeezed observations made the visualization hard to read, hence, countries with significant outliers were selected for visualization.

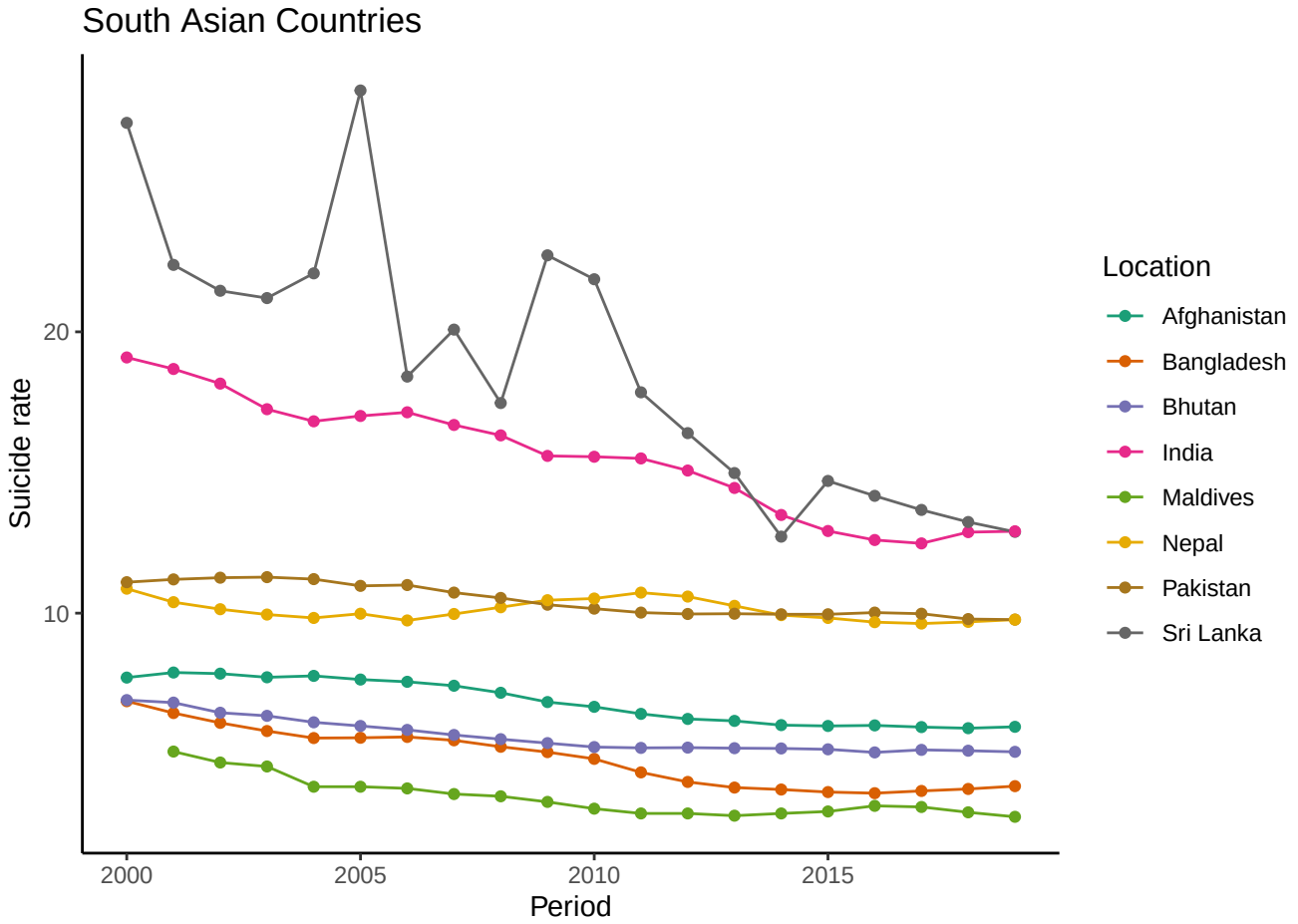


Visualization 3: Trend of suicide rates in South Asian countries. This could be paired with any of the article with a mention of a South Asian country, for example, Malaysia in

Suicide is not a crime, Pakistan in *With suicide not a crime, the real work begins*, Bangladesh in *Suicide is a mental health issue, not a crime* and a discussion on *India's Mental Health act*. Sri Lanka is also mentioned in The alternatives that can help prevent suicide.

Data source: WHO

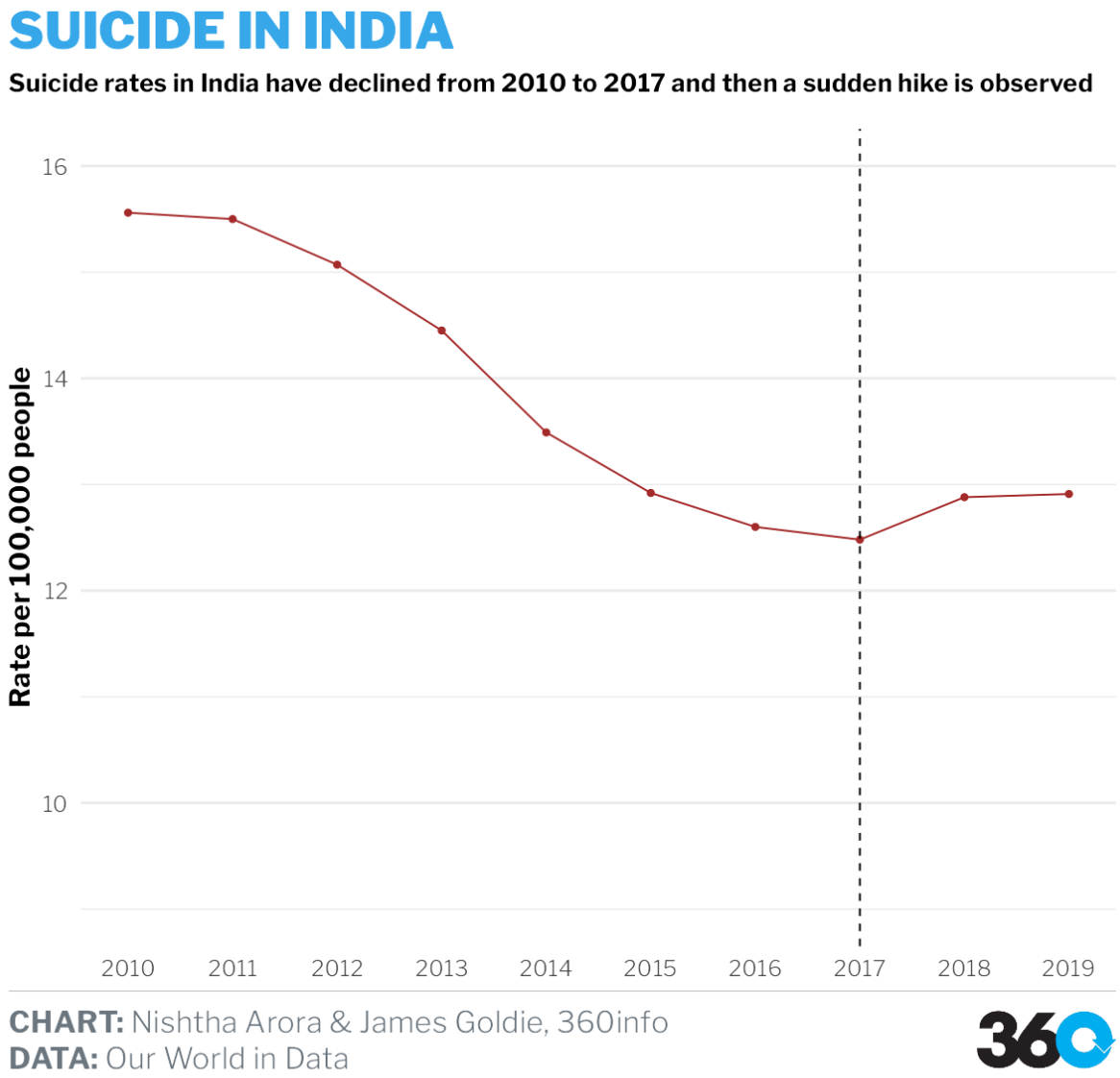
The suicide rates in South Asian countries are reported to be between 0.43 to 331.0 per 100,000 population, which is high compared to the world average.



Shortlisted Visualizations

Visualization 4: A time-series plot depicting suicide rate trend before and after 2017, i.e. to pair up with the article on India's Mental Health Act 2017.

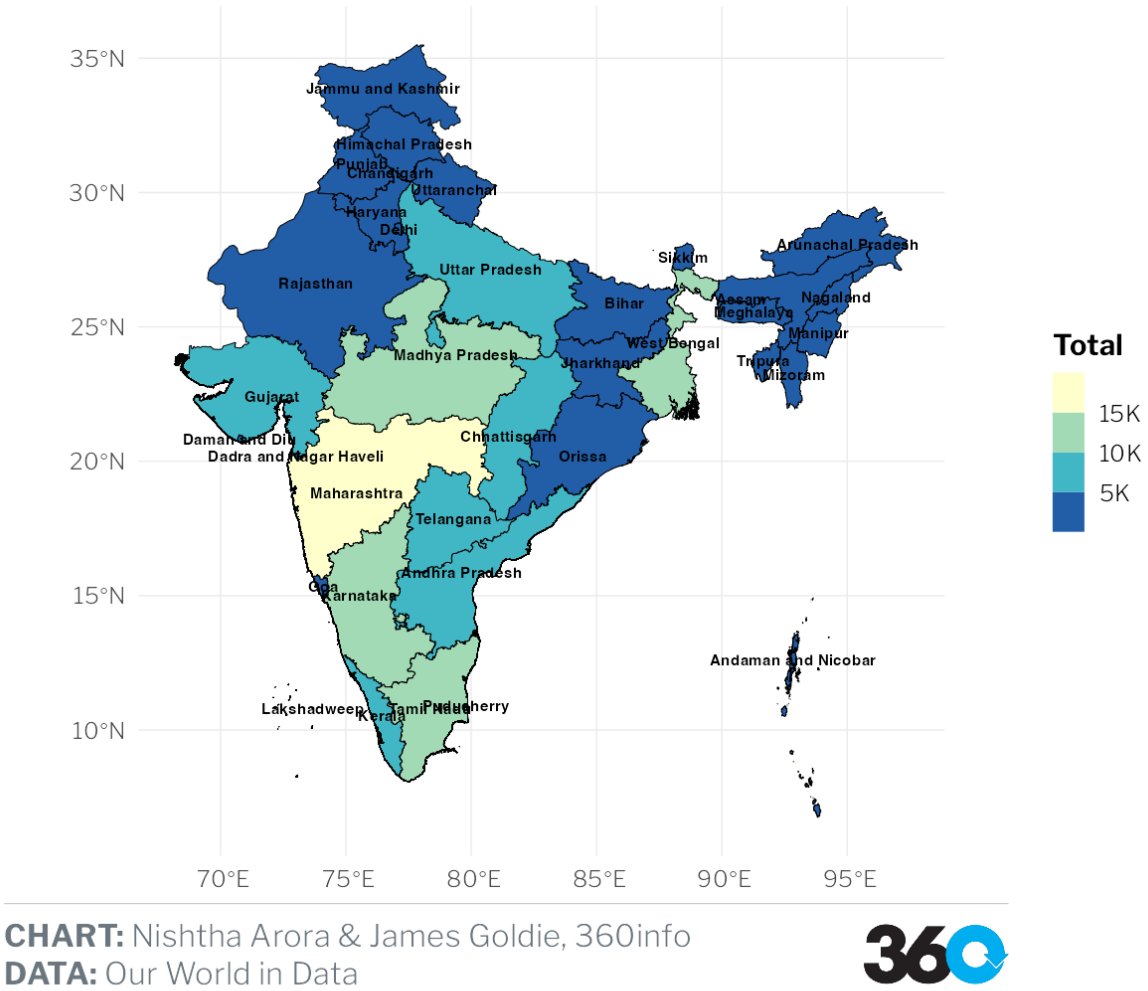
Data source : WHO



Visualization 5: A state wise India's suicide rate to pair with the article How India contunues to punish those who attempt suicide..

Data source:

- Data.gov
- Geometery data at Diva-gis



Selected Visualization

Visualization 6: Comparison of visualization 4 with the similar plot made from National Crime Records Bureau extracted data

Data source:

- NCRB

```
## mapping: x = ~x, y = ~y
## geom_rich_text: na.rm = FALSE, label.r = 0, label.padding = 0.25
## stat_identity: na.rm = FALSE
## position_identity
```

SUICIDE IN INDIA (data.gov)

Suicide rates in India have rapidly fallen till 2016.

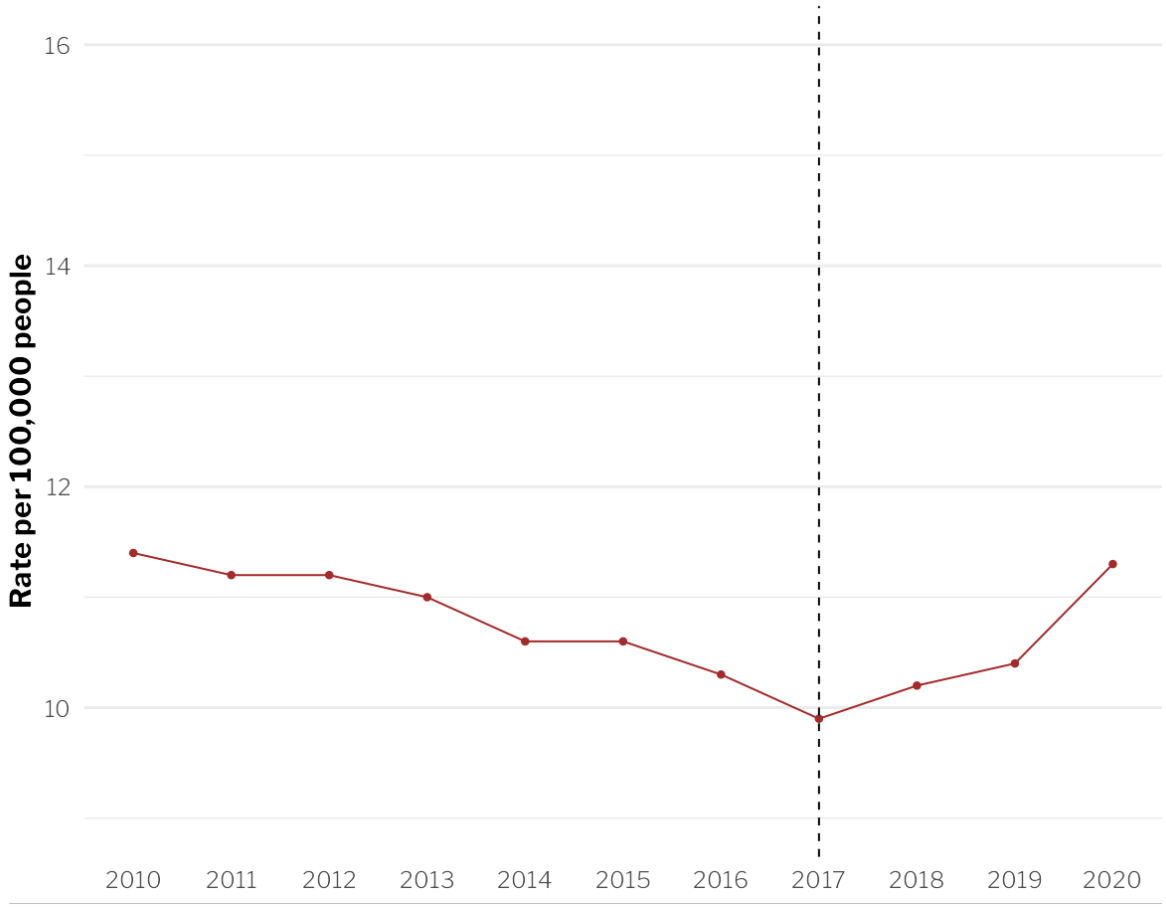


CHART: Nishtha Arora & James Goldie, 360info
DATA: NCRB

Policing the Police

The visualizations here are generic and related to different aspects of 'Policing the Police', for example,

- Corruption/bribery
- Police Shootings/Encounter/Drug Wars/ Causalities
- Trust in Police

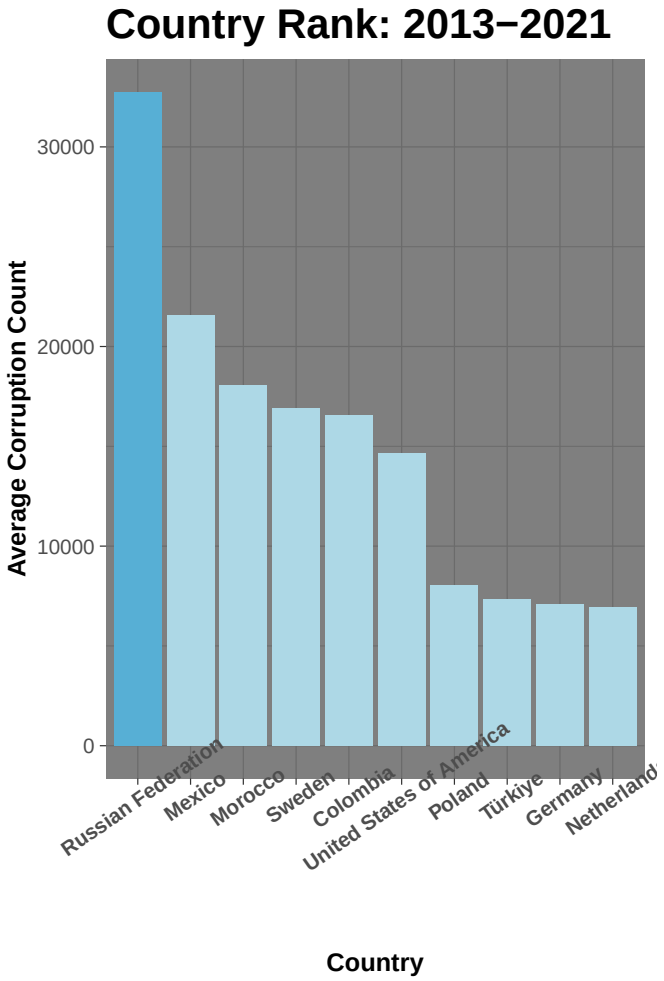
This is because, the package publication date was in June, i.e. 2-3 weeks later than the last day of internship. The article drafts would be created in early June and hence, the visualizations are not paired to specific articles for this package.

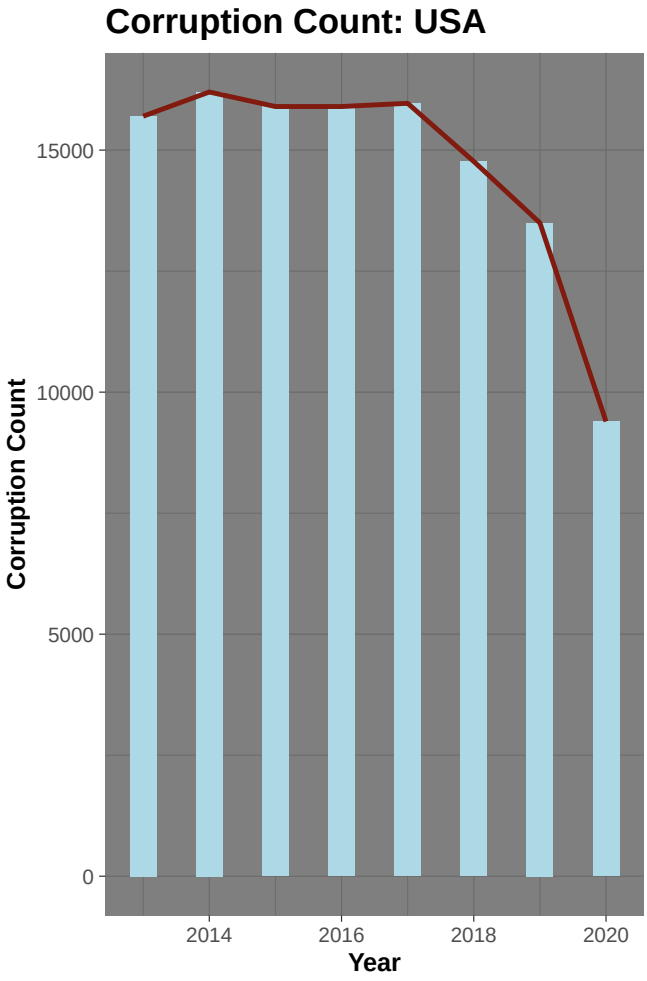
Initial Visualizations

Visualizations 1 & 2

Data source:

- UNODC

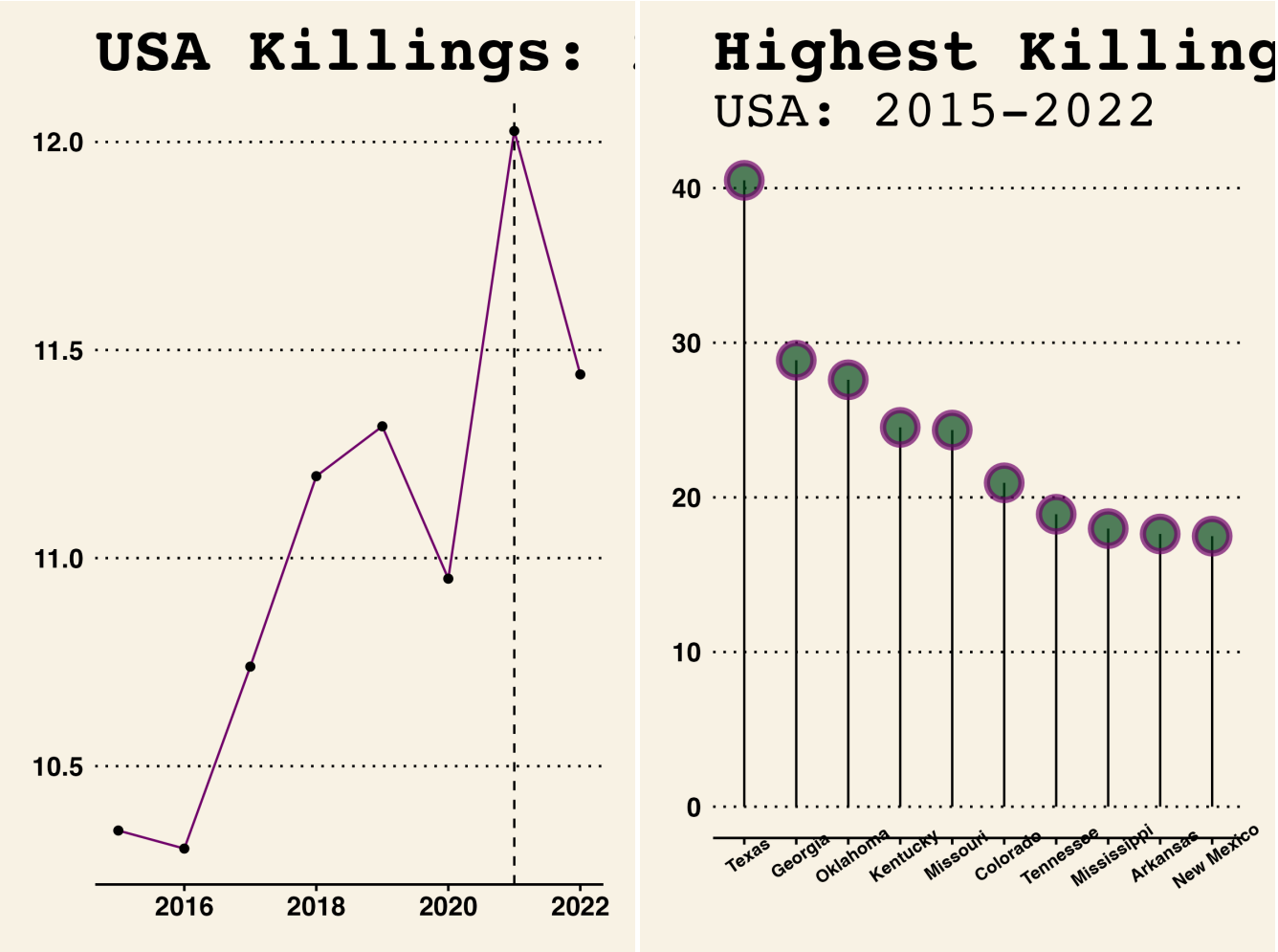




Visualization 3 & 4

Data Sources:

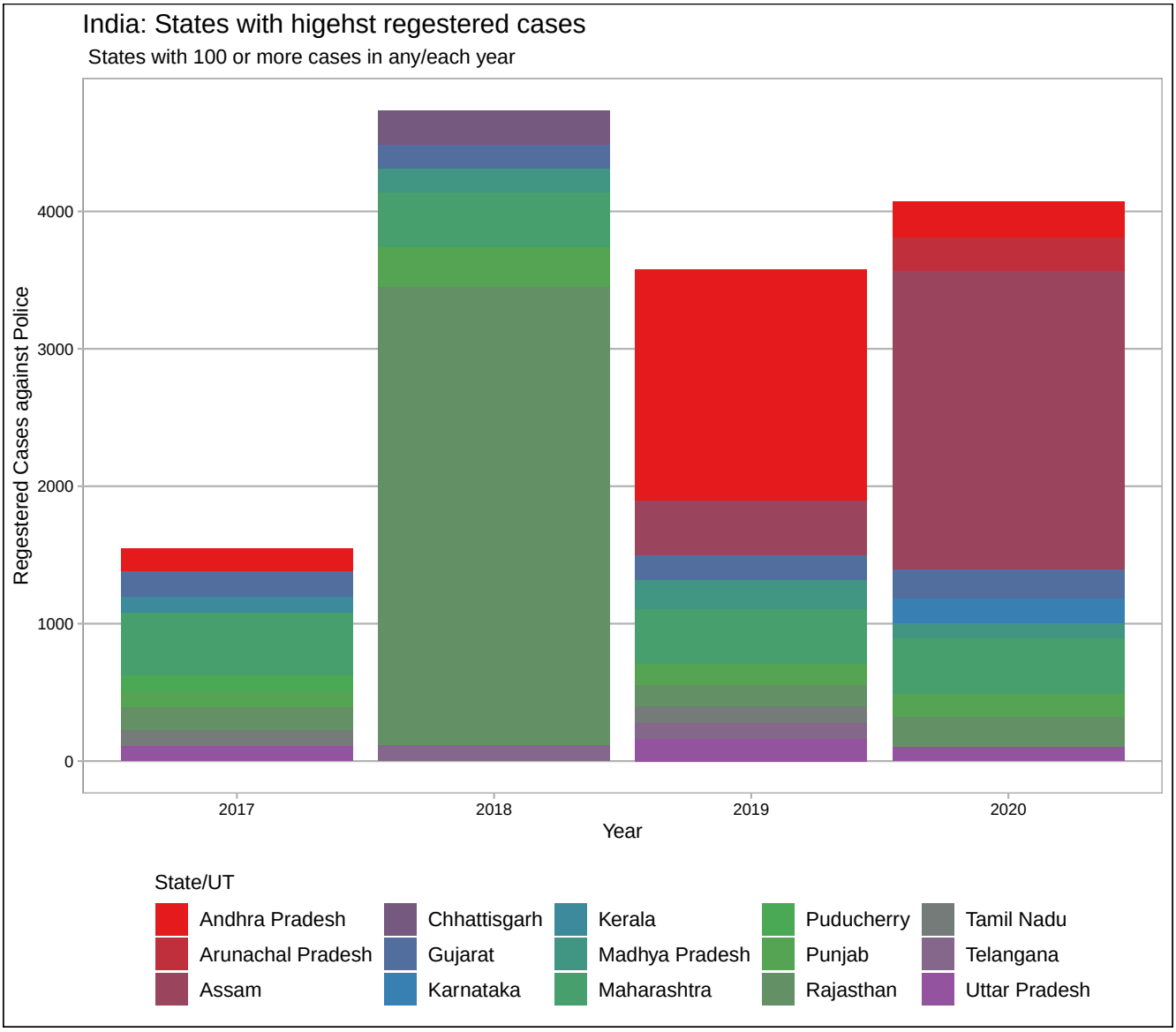
- Washinton Post
- Washinton Post Github
- Census data 2010-2020
- Census data 2020-2022
- Stats/States code data github

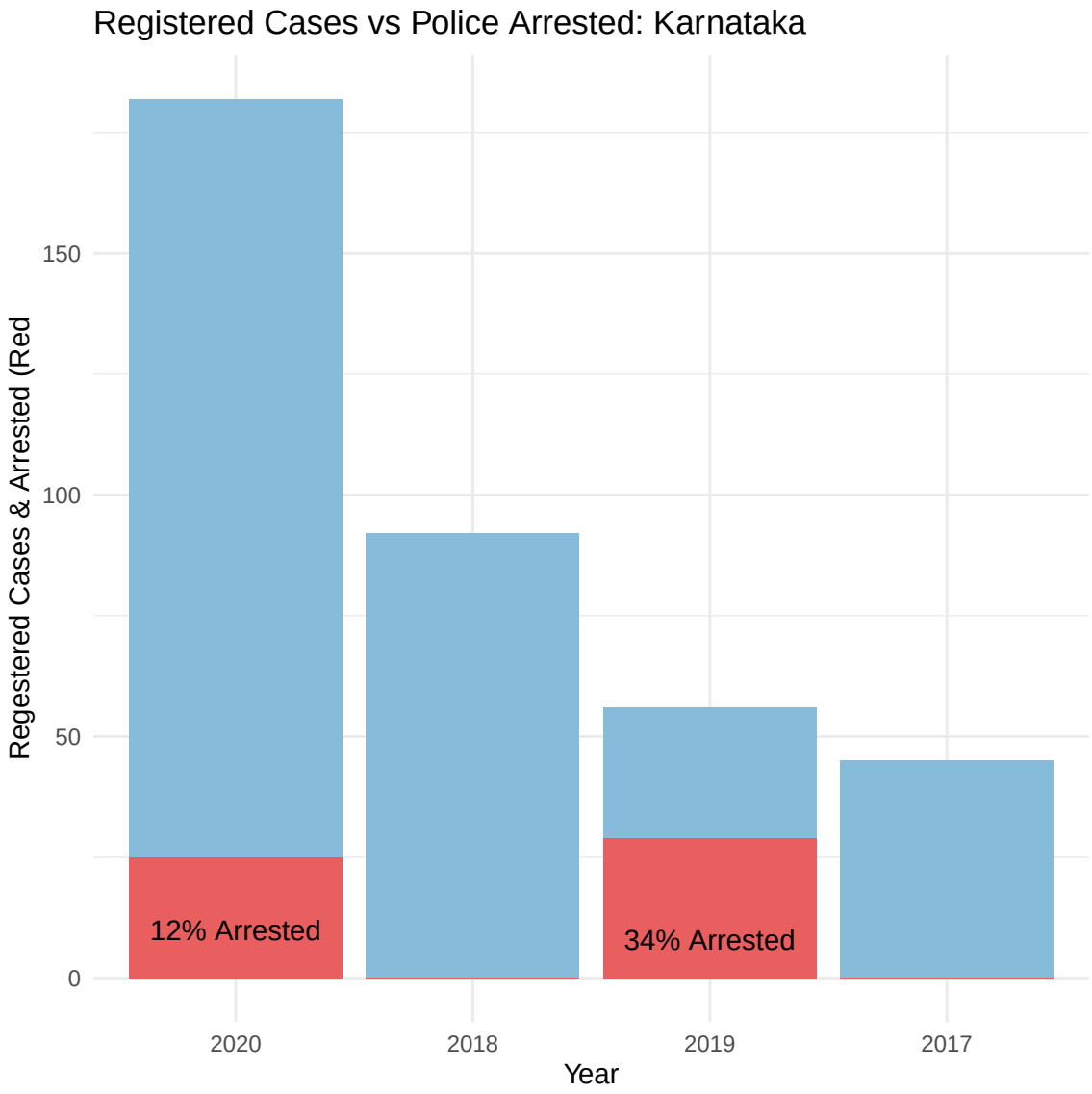


Visualizatio 5 & 6

Data Source:

- NCRB

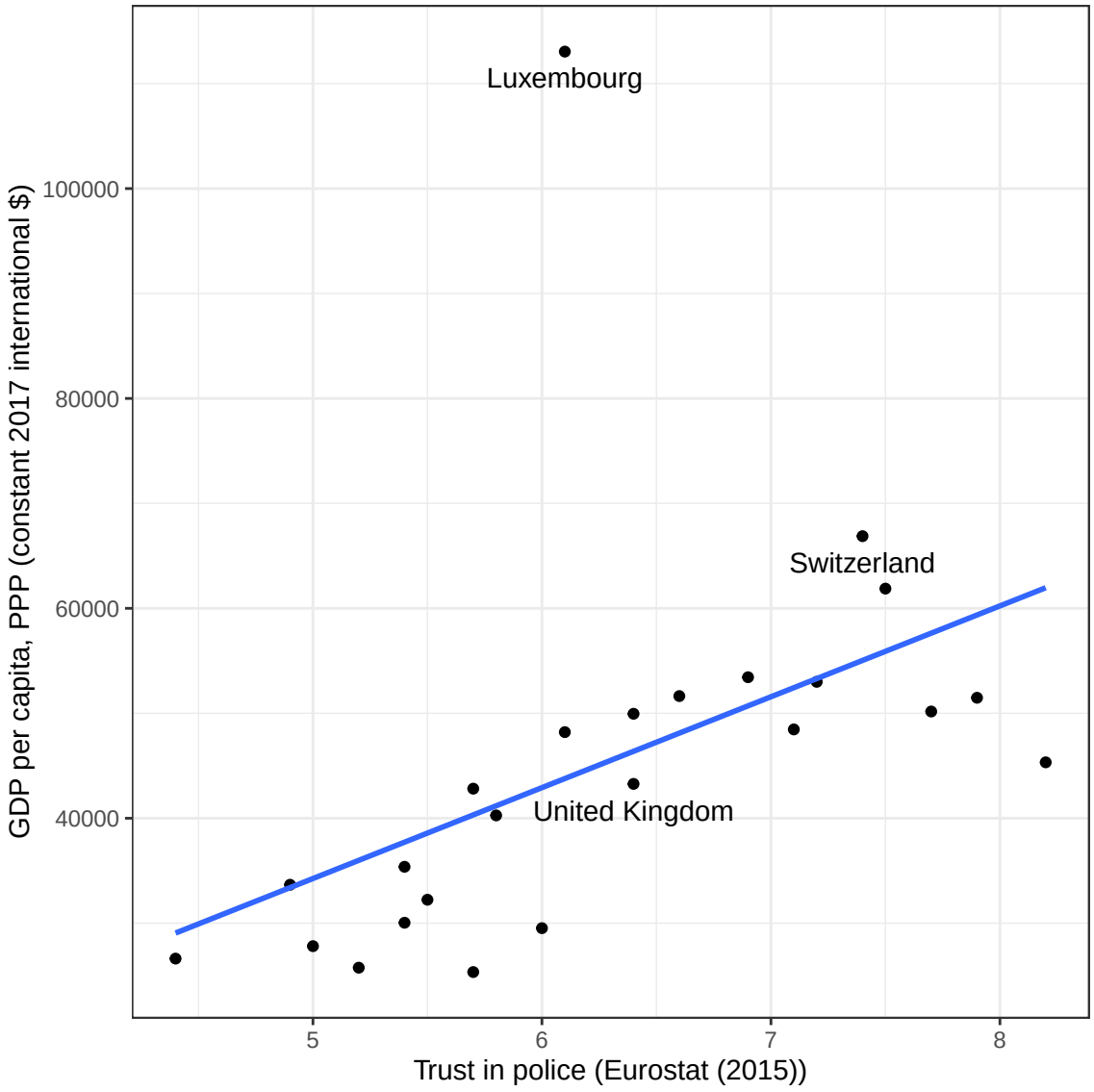

India: States with higehst regestered cases
States with 100 or more cases in any/each year
Regestered Cases against Police
0
1000
2000
3000
4000
2017
2018
2019
2020
Year
State/UT
Andhra Pradesh
Arunachal Pradesh
Assam
Chhattisgarh
Gujarat
Karnataka
Kerala
Madhya Pradesh
Maharashtra
Puducherry
Punjab
Rajasthan
Tamil Nadu
Telangana
Uttar Pradesh



Visualization 7

Data Source:

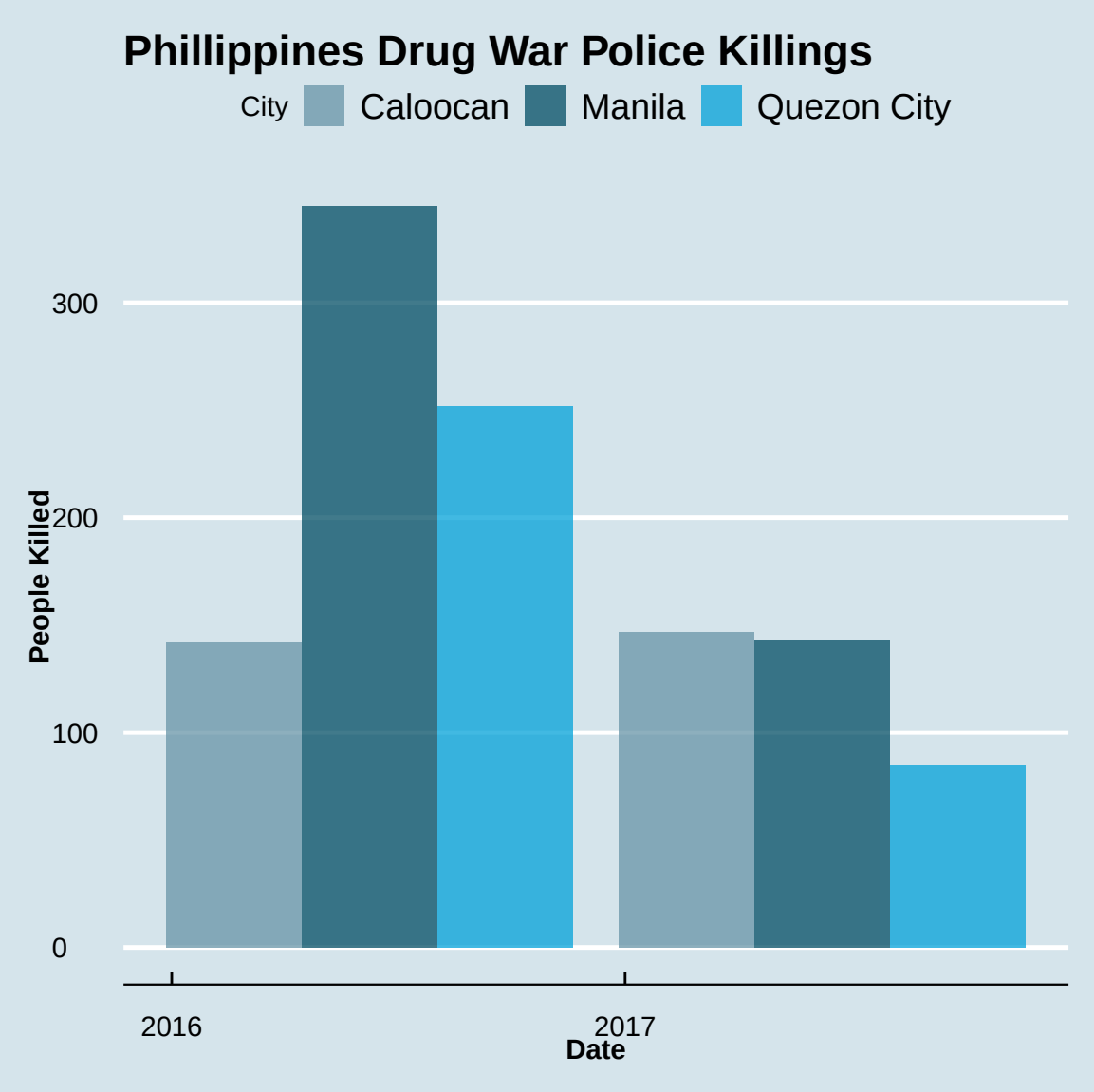
- OWID



Visualization 8

Data Source:

- Data.World

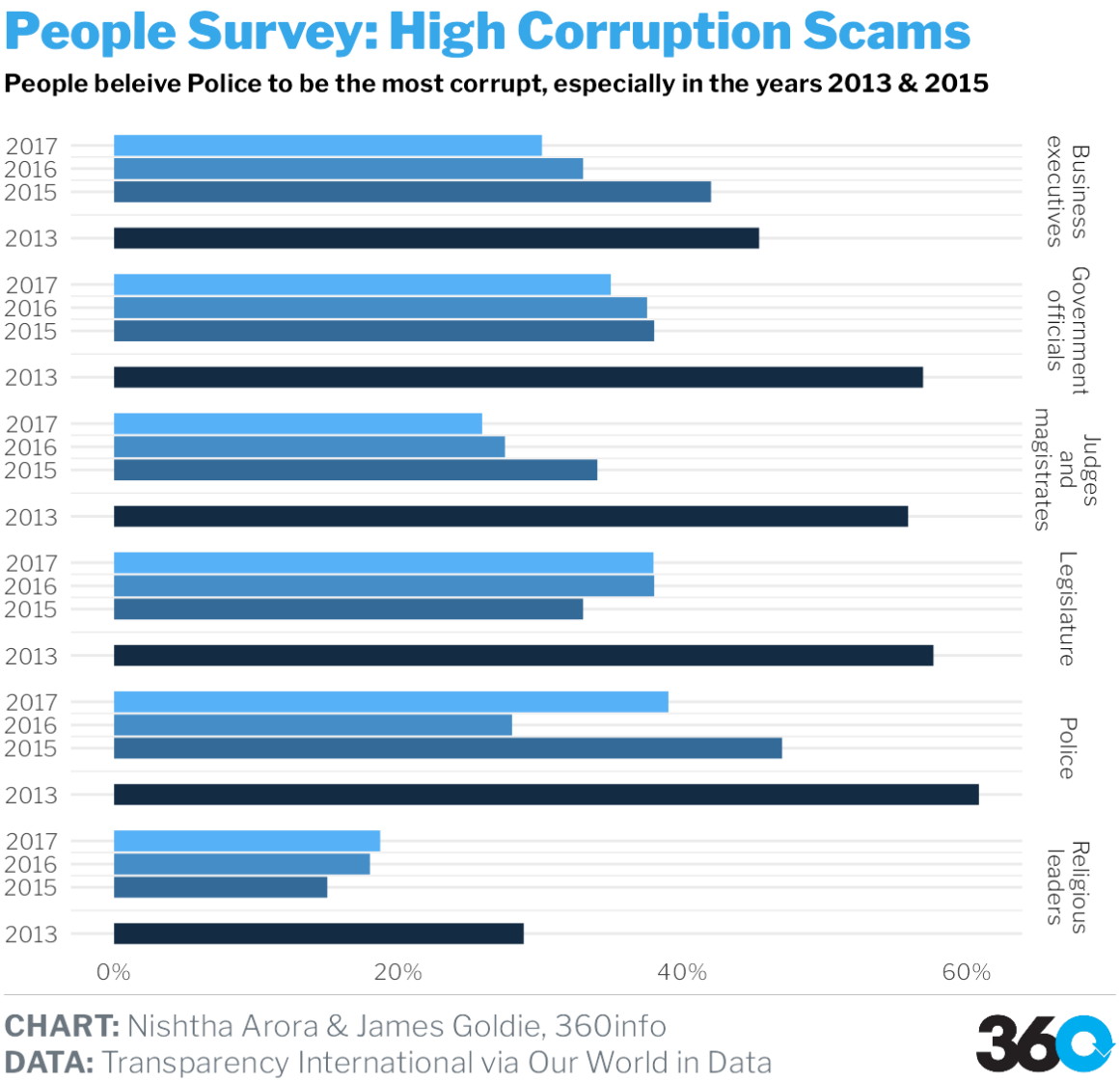


Short- Listed Visualizations

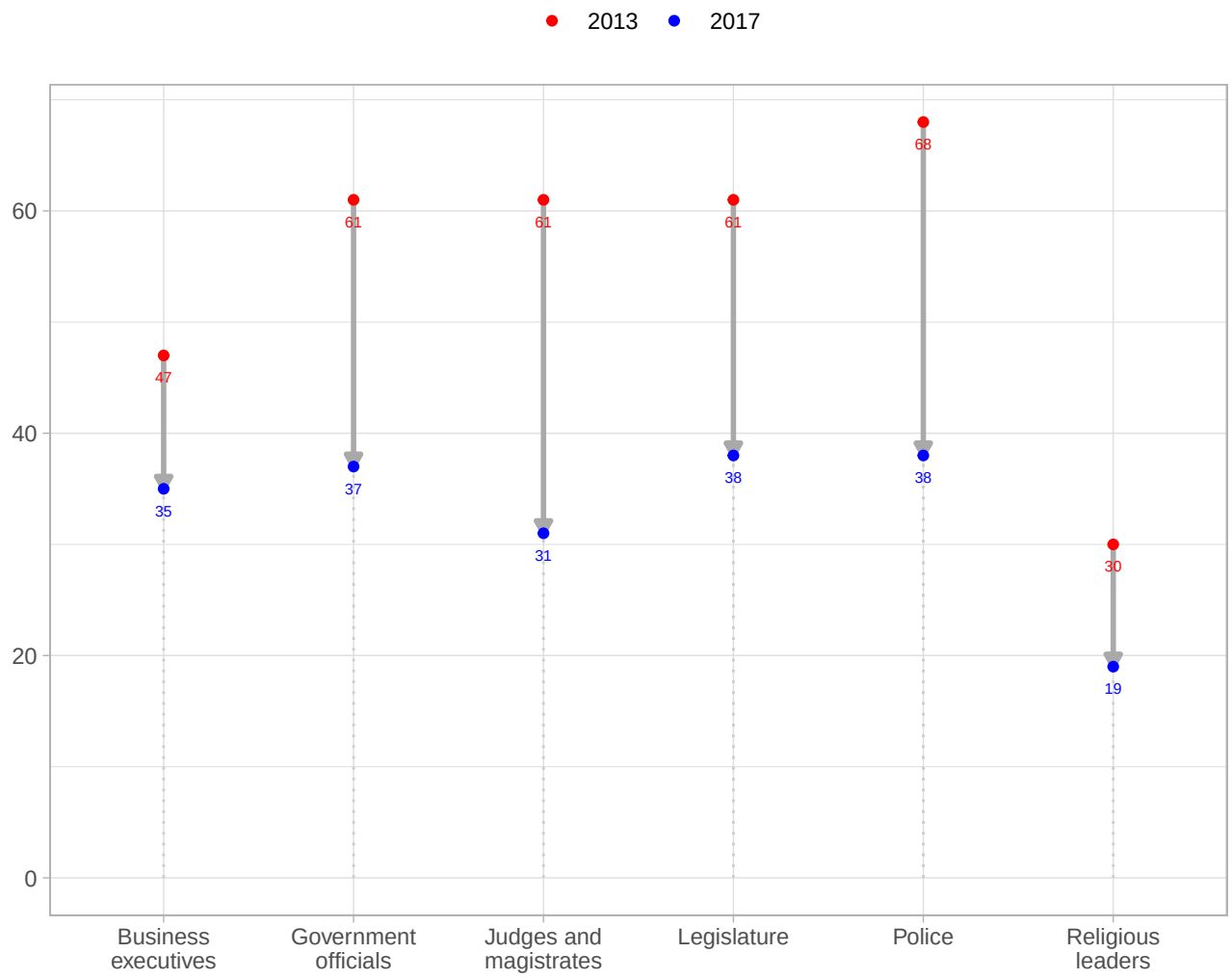
Visualization 9

Data Source:

- OWID

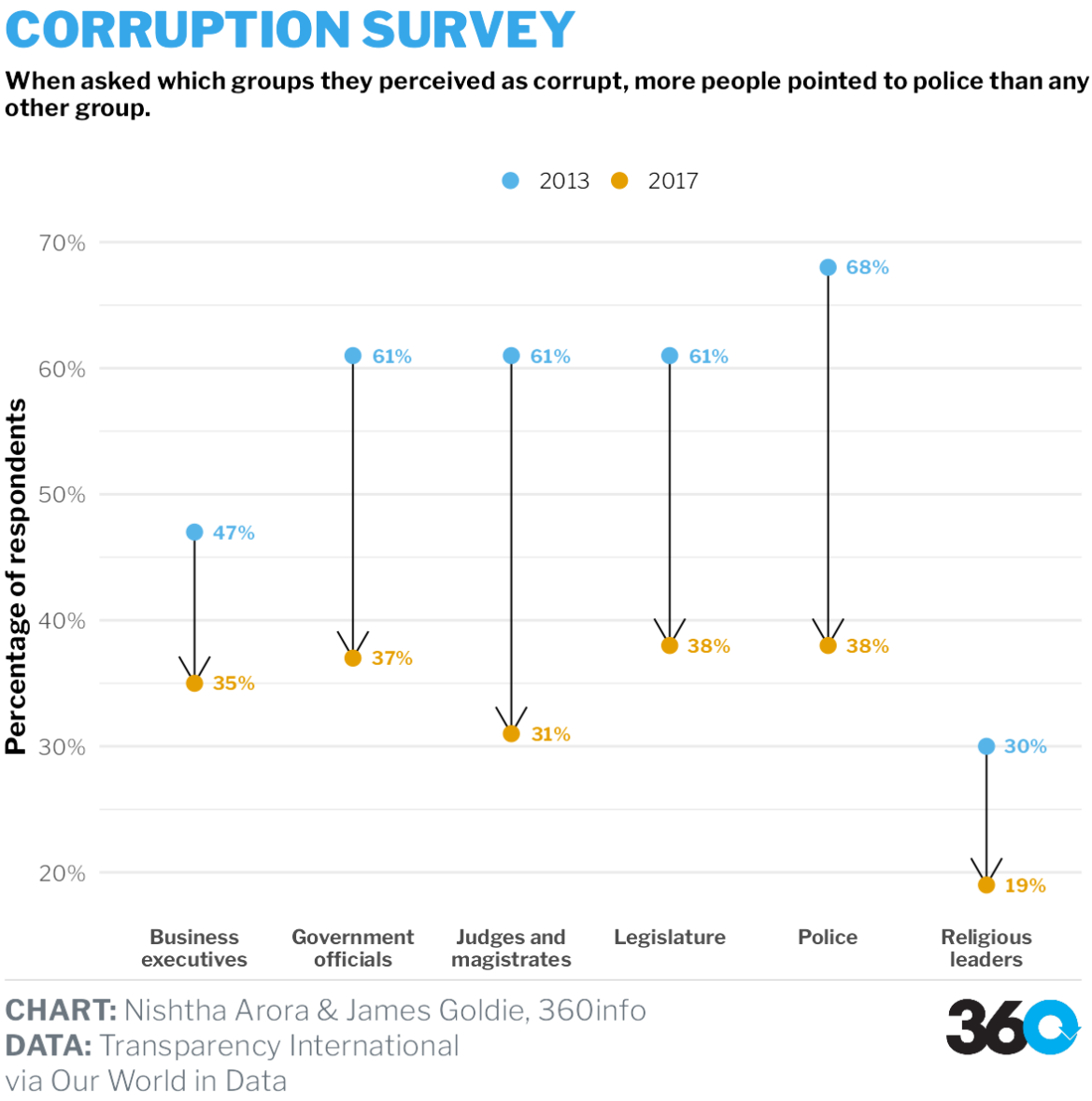

People Survey: High Corruption Scams
People beleive Police to be the most corrupt, especially in the years 2013 & 2015
2017
2016
2015
2013
Business executives
Government officials
Judges and magistrates
Legislature
Police
Religious leaders
0%
20%
40%
60%
CHART: Nishtha Arora & James Goldie, 360info
DATA: Transparency International via Our World in Data

Visualization 10 - updating visualization 9



Selected Visualization

Visualization 11- updating visualization 10



Conclusion

References

360info. (n.d.). https://360info.org

Assisted dying - Christianity. (n.d.). Christianity. https://christianity.org.uk/article/assisted-dying

Basic question for plotting x axis using row.names. (n.d.). Google Groups. https://groups.google.com/g/ggplot2/c/UkmDYDcRNWc?pli=1

Chang, W. (2023, May 28). 7.1 Adding Text Annotations | R Graphics Cookbook, 2nd edition. https://r-graphics.org/recipe-annotate-text

Combine two data frames with the same column names. (n.d.). Stack Overflow. https://stackoverflow.com/questions/20081256/combine-two-data-frames-with-the-same-column-names

Crime in India Table Contents | National Crime Records Bureau. (n.d.). https://ncrb.gov.in/en/crime-in-india-table-addtional-table-and-chapter-contents?field_date_value%5Bvalue%5D%5Byear%5D=&field_select_table_title_of_crim_value=18&items_per_page=All

Dattani, S. (2023, April 2). Suicides. Our World in Data. https://ourworldindata.org/suicide

Formatting Decimal places in R. (n.d.). Stack Overflow. https://stackoverflow.com/questions/3443687/formatting-decimal-places-in-r

G, E. (1987). Hawkins, D. M.: Identification of Outliers. Chapman and Hall, London – New York 1980, 188 S., £ 14, 50. Biometrical Journal, 29(2), 198. https://doi.org/10.1002/bimj.4710290215

GeeksforGeeks. (2021). Change column name of a given DataFrame in R. GeeksforGeeks. https://www.geeksforgeeks.org/change-column-name-of-a-given-dataframe-in-r/

ggplot2 axis scales and transformations - Easy Guides - Wiki - STHDA. (n.d.). http://www.sthda.com/english/wiki/ggplot2-axis-scales-and-transformations

ggplot2 scatter plots : Quick start guide - R software and data visualization - Easy Guides - Wiki - STHDA. (n.d.). http://www.sthda.com/english/wiki/ggplot2-scatter-plots-quick-start-guide-r-software-and-data-visualization

Ghana - Criminal Code 1960 (Act 29). (n.d.). https://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=88530

Halevy, A., & McGregor, S. E. (2012). Data Management for Journalism. IEEE Data(Base) Engineering Bulletin, 35, 7–15. http://sites.computer.org/debull/A12sept/p7.pdf

Highlight a single “bar” in ggplot. (n.d.). Stack Overflow. https://stackoverflow.com/questions/45820250/highlight-a-single-bar-in-ggplot

how to add lines over a column bar graph where the lines pass by the middle-top of the bars considering bars with position='dodge'? (n.d.). Stack Overflow. https://stackoverflow.com/questions/

72116660/how-to-add-lines-over-a-column-bar-graph-where-the-lines-pass-by-the-middle-top

how to wrap text in ggplot for facet_grid labels. (n.d.). Stack Overflow. https://stackoverflow.com/questions/43796409/how-to-wrap-text-in-ggplot-for-facet-grid-labels

Jordans, M. J. D., Kaufman, A., Brenman, N. F., Adhikari, R. K., Kohrt, B. A., Tol, W. A., & Komproe, I. H. (2014). Suicide in South Asia: a scoping review. BMC Psychiatry, 14(1). https://doi.org/10.1186/s12888-014-0358-9

Kanevsky, G. (2013). How to expand color palette with ggplot and RColorBrewer | R-bloggers. R-bloggers. https://www.r-bloggers.com/2013/09/how-to-expand-color-palette-with-ggplot-and-rcolorbrewer/

Kenya - The Penal Code (Cap. 63). (n.d.). https://www.ilo.org/dyn/natlex/natlex4.detail?p_isn=28595&p_lang=en

Lester, D. (2006). Suicide and Islam. Archives of Suicide Research, 10(1), 77–97. https://doi.org/10.1080/13811110500318489

Lua filters in R Markdown. (n.d.). https://rmarkdown.rstudio.com/docs/articles/lua-filters.html

Ochuku, B. K., Johnson, N. M., Osborn, T. L., Wasanga, C., & Ndetei, D. M. (2022). Centering decriminalization of suicide in low – and middle – income countries on effective suicide prevention strategies. Frontiers in Psychiatry, 13. https://doi.org/10.3389/fpsyt.2022.1034206

Osborne, J. A., & Overbay, A. (2004). The power of outliers (and why researchers should ALWAYS check for them). Practical Assessment, Research and Evaluation, 9(1), 1–8. https://doi.org/10.7275/qf69-7k43

Muiruri, P. (2022, October 19). Concern grows in Kenya after alarming rise in suicide cases. The Guardian. https://www.theguardian.com/global-development/2021/aug/10/concern-grows-in-kenya-after-alarming-rise-in-suicide-cases

Plot data in descending order as appears in data frame. (n.d.). Stack Overflow. https://stackoverflow.com/questions/16961921/plot-data-in-descending-order-as-appears-in-data-frame

Ranjan, R., Kumar, S., Pattanayak, R. D., Dhawan, A., & Sagar, R. (2014). (De-) criminalization of attempted suicide in India: A review. Industrial Psychiatry Journal, 23(1), 4. https://doi.org/10.4103/0972-6748.144936

Riederer, Y. X. C. D. E. (2022, November 7). 10.1 The function knitr::kable() | R Markdown Cookbook. https://bookdown.org/yihui/rmarkdown-cookbook/kable.html

Side By Side Bar Graphs In R & ggplot2. (n.d.). https://dk81.github.io/dkmathstats_site/rvisual-sidebyside-bar.html

Suicide Decriminalisation - United for Global Mental Health. (2022, July 4). United for Global Mental Health. https://unitedgmh.org/knowledge-hub/suicide-decriminalisation/?utm_campaign=SuicideDecrimReport&utm_medium=referral&utm_source=vip&utm_content=SuicideDecrimReport

Sum across multiple columns with dplyr. (n.d.). Stack Overflow. https://stackoverflow.com/questions/28873057/sum-across-multiple-columns-with-dplyr

World Health Organization. (2021). Comprehensive mental health action plan 2013–2030. https://apps.who.int/iris/handle/10665/345301

World Health Organization: WHO. (2021). Suicide. www.who.int. https://www.who.int/news-room/fact-sheets/detail/suicide

Zhu, H. Z. (2021). Create Awesome LaTeX Table with knitr::kable and kableExtra. cran.r-project.org. Retrieved May 10, 2023, from https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_pdf.pdf