

onemorelight()

making better R functions

Yaning Wu

Outline

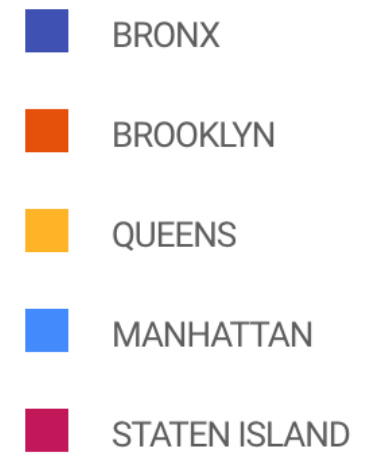
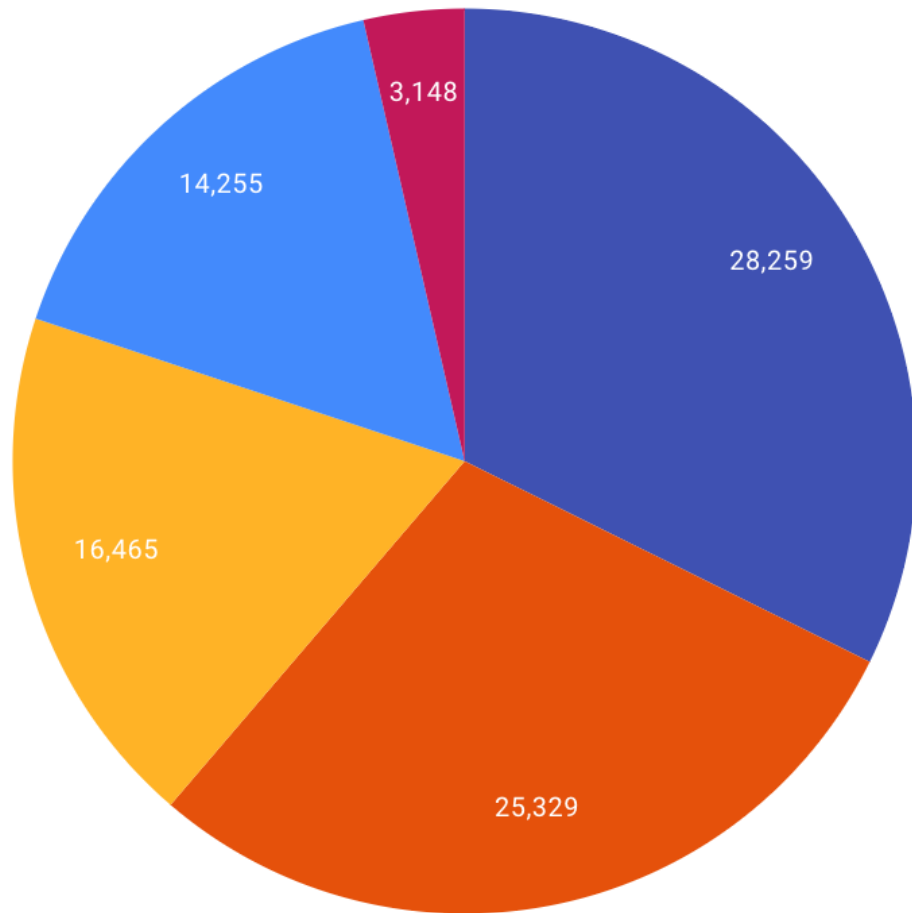
- Part 1: How do we tend to visualise big numbers?
- Part 2: One stand-out example
- Part 3: My first function
- Part 4: Making a better function
- Part 5: Applying the function to example data
- Part 6: Reflecting on the process

Part 1: Visualising mass happenings

TW: distressing themes

Sometimes, big things happen to many

- Conflicts, epidemics, large-scale events, climate change ...
- How do we describe these happenings with clarity and care?



Evictions in New York City

Credit:

NYC Open Data

Three statements

Child mortality throughout history

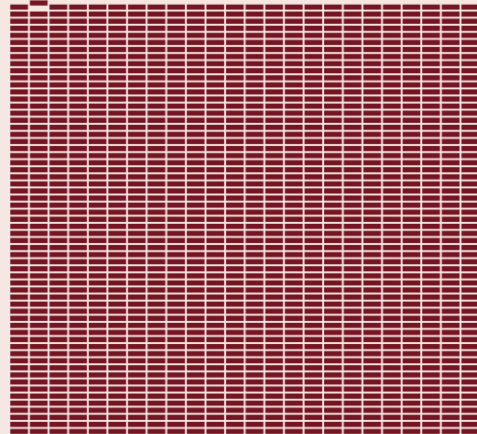
All three statements are true at the same time

Our World
in Data

The world is awful

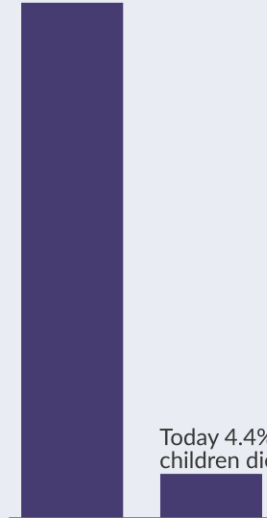
Globally 4.3% of all children die before they are 15 years old.
This means every day 16,000 children die.

Each rectangle represents the death of 11 children.
That's the number of children who die every minute.



The world is much better

In the past around 50% of all children died



Today 4.4% of all children die

The world can be much better

4.4% of all children worldwide die



In the European Union 0.47% of all children die.

→ We know that it is possible to make the world a better place.

Data: 2020 data from UN IGME and historical data from Volk and Atkinson (2013).
[OurWorldinData.org](https://ourworldindata.org) - Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Max Roser

Credit:
Our World in Data



Woman, life, freedom

Protesters killed in Iran

Credit:

[@federica.fragapane](#) (Instagram)

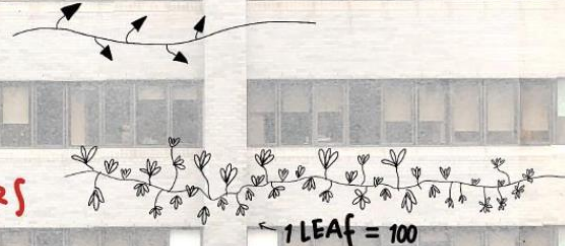
INCREASE IN VOLUNTEERING, NYC

500 PROJECTS

4,500 VOLUNTEERS

15,000 HOURS
of VOLUNTEERING

450,000 MEALS
HAVE BEEN DISTRIBUTED

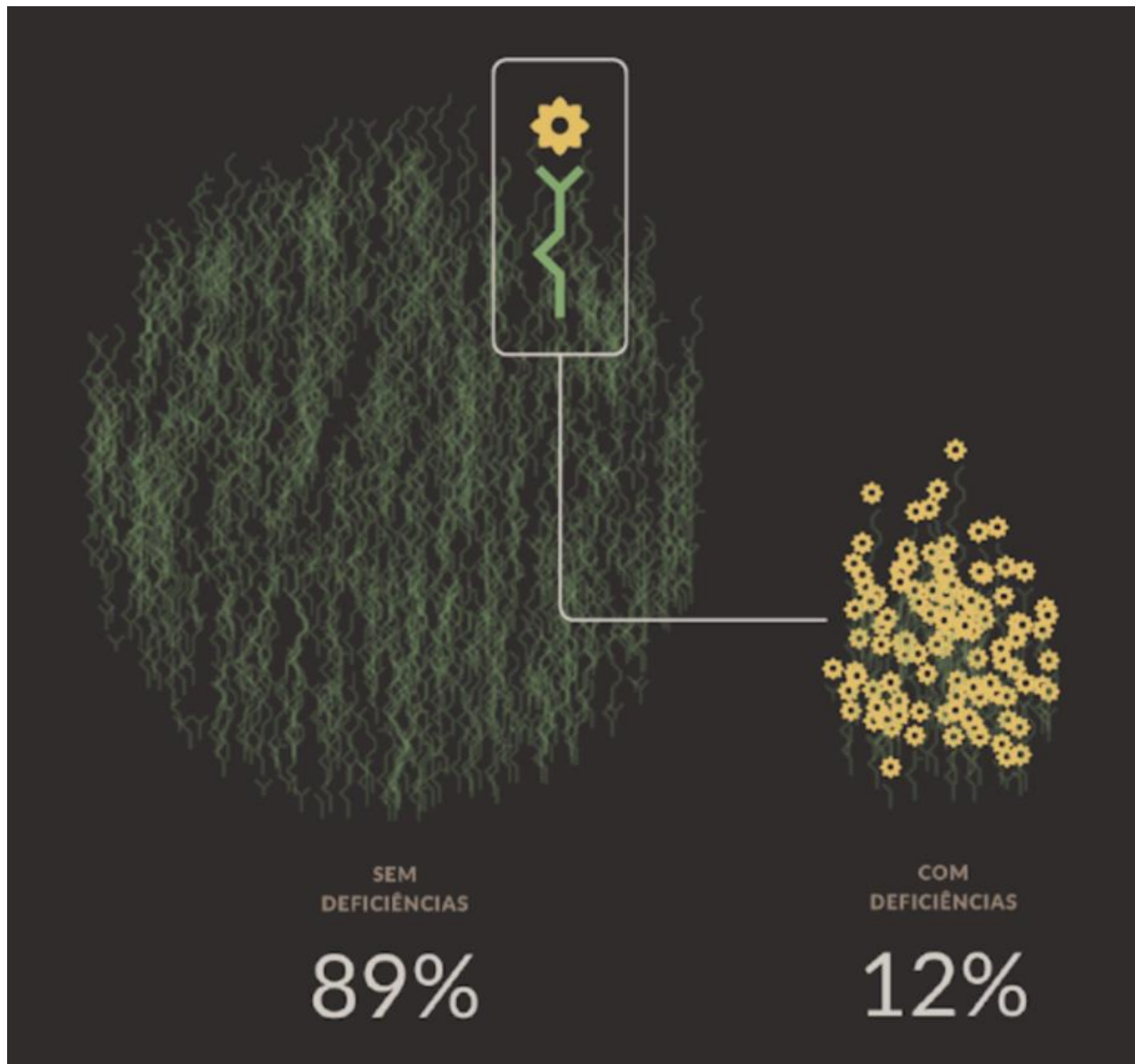


Community

Volunteering activity during the pandemic

Credit:

@giorgialupi (Instagram)

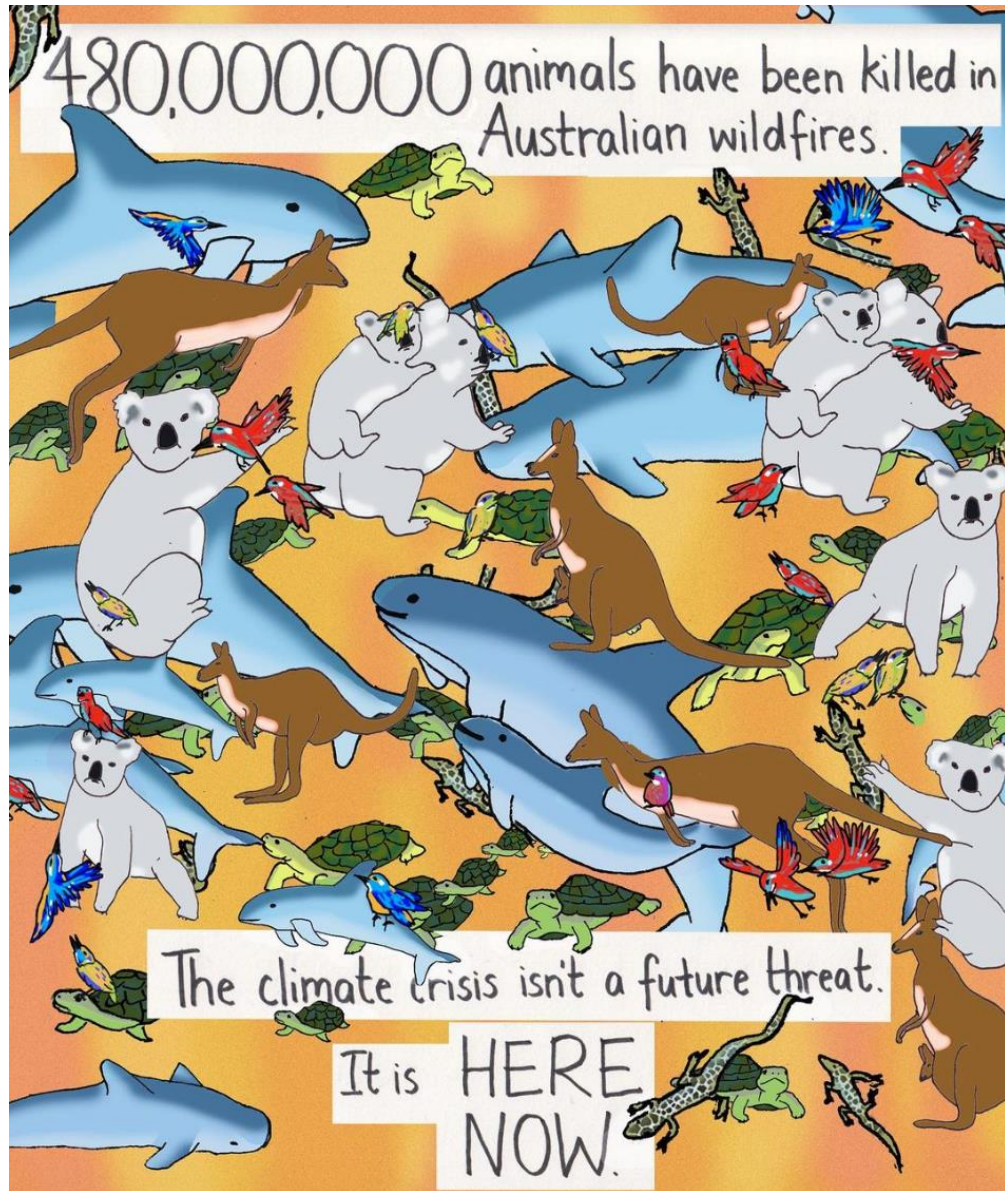


Sunflowers

Adoption in Brasil

Credit:

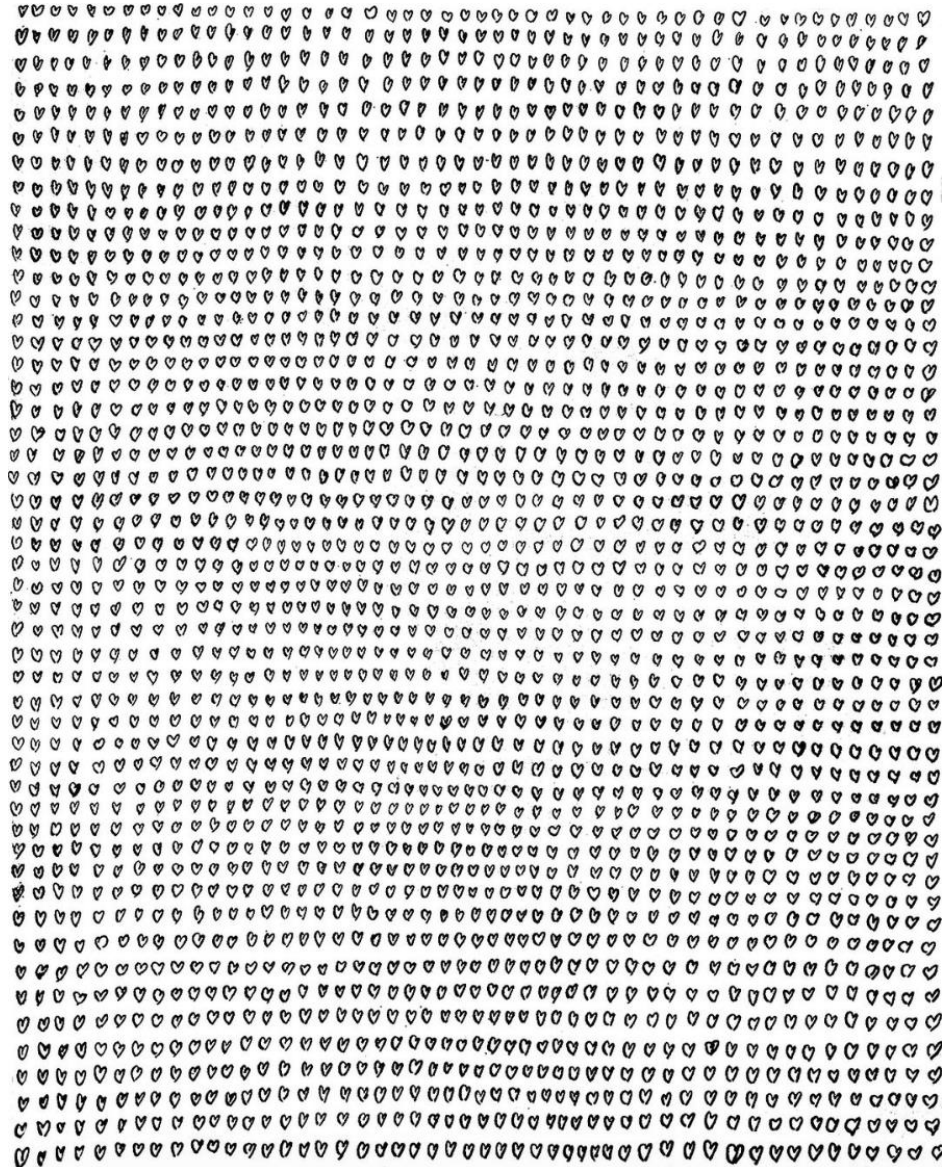
Estadão (web)



Self-explanatory

Credit:

@monachalabi (Instagram)



So many hearts

People killed in Gaza

Credit:

[@so.many.hearts](#) (Instagram)

Do you notice contrasting approaches?

- Personal → Detached
- Granular → Zoomed out
- Variable → Uniform

What works best?

Depends on your purpose ...

Part 2: The New York Times' "**Wall of Grief**"

"All the News That's Fit to Print"

The New York Times

Late Edition
Today: partly cloudy, high 58, low 38. Tomorrow: overcast, high 58, low 38. Weather map on page 10.

VOL. CLXX... No. 58,976

NEW YORK, SUNDAY, FEBRUARY 21, 2021

\$6.00

U.S. VIRUS DEATHS NEARING 500,000 IN JUST ONE YEAR

MORE THAN IN 3 WARS

Empty Spaces in Cities, Towns, Restaurants, Homes and Hotels

By JULIE BOSMAN
CHICAGO — A nation numbed by misery and loss is confronting a number that still has the power to shock: 500,000.

Roughly one year since the first known death by the coronavirus in the United States, an astonishing toll is nearing — the loss of half a million people.

No other country has endured so many deaths in the pandemic. More Americans have perished from COVID-19 than died in World War I and the Vietnam War combined.

The toll is a stark reminder of the human cost of the virus. New virus cases are down, but deaths are rising and vaccines are steadily being administered.

But there is concern about emerging variants, the virus's persistence in communities and the impact of mask-wearing. Each death has left an empty space in a community, a regular need to sit on the side of a road, a hole in a family's fabric without a crack.

Russia Fears But Can't Quit Open Internet

By ANTON TROIANOVSKI
MOSCOW — Margaret Thatcher, the prime minister of the United Kingdom, once said that the Internet is the greatest invention since the printing press.

But in Russia, the government is wary of the Internet's power to spread dissent. It has cracked down on social networks and messaging apps, but it cannot afford to shut them down completely.

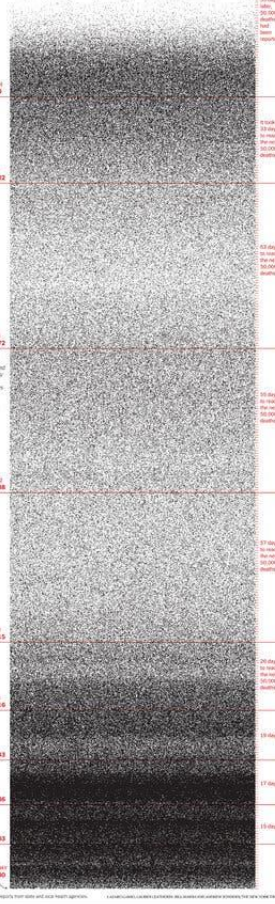
While the Russian Internet is open, it is heavily censored. The government has blocked access to many Western websites and has created a parallel, state-controlled network.

But even as Mr. Putin faces the digital divide, his government is pushing to create a "sovereign" Internet. This would allow the government to control the flow of information within its borders.

The Toll: America Approaches Half a Million Covid Deaths

Feb. 20, 2021. First report of a U.S. death in Washington State.

Each dot represents one death from Covid-19 in the U.S.



Garland Faces Resurgent Peril Of Extremism

Oklahoma City Attack Shaped His Views

By MARK LEBROWICZ
WASHINGTON — Judge Mark E. Garland always made a point of wearing a coat and tie when he transported the wreckage of an aircraft that had crashed after Thursday 9/11.

He had been dispatched from Washington to oversee the case for the Justice Department, and he said colleagues that he viewed the daily routine as a statement of respect for an event that had shaken the nation.

But Judge Garland's views on extremism have become more pronounced since the Oklahoma City bombing. He has said that the attack was a "wake-up call" for the nation and that it has shaped his views on the rule of law.

Garland has been a vocal critic of the far-right movement and has said that he is concerned about the rise of extremism in the United States. He has also said that he is committed to protecting the Constitution and the rule of law.

Lack of Birth Control Deepens Women's Burden in Venezuela

By JULIE YERKOWITZ and HEAVEN HERBERG
SAN JUAN DE LOS RIOS, VENEZUELA — The Internet is the greatest invention since the printing press.

But in Venezuela, the government is wary of the Internet's power to spread dissent. It has cracked down on social networks and messaging apps, but it cannot afford to shut them down completely.

While the Venezuelan Internet is open, it is heavily censored. The government has blocked access to many Western websites and has created a parallel, state-controlled network.

But even as Mr. Maduro faces the digital divide, his government is pushing to create a "sovereign" Internet. This would allow the government to control the flow of information within its borders.

STORMS EXPOSING A NATION PRIMED FOR CATASTROPHE

CLIMATE CHANGE WRATH

Unprepared for Threats Facing Power Grids, Water and Roads

By CHRISTOPHER PLANTE, BRAD FISHER and HELEN BEHRE
Even as Texas struggled to restore electricity and water over the past week, signs of the risks posed by increasingly extreme weather in America using old infrastructure were cropping up across the country.

The storm opening winter storm triggered blackouts in Texas, Oklahoma, Missouri and several other states. One-third of all production in the nation was halted. Drinking water systems in some were disrupted.

Road networks nationwide were closed and millions of vehicles in 20 states were disrupted. The crisis exposes a profound warning: As climate change brings more frequent and intense storms, floods, heat waves, wildfires and other extreme events, it is a growing problem across the length of the country's energy, water, power, gas, and transportation infrastructure.



Juliana Guzman with two of her six children in northern Venezuela, where contraptions are difficult to find, as shown above.

TRACKING AN OUTBREAK: 4
Israel Finds Vaccines for Syria
To secure the release of an Israeli hostage, Israel secretly agreed to transport a supply of Russian-made COVID-19 vaccines for Syria, an official said. PAGE 1

SPORTS 23-37
Oaks Win Australian Open
Naomi Osaka, 23, won a 1-6, 4-6 Grand Slam tennis title after a straight-set victory over Jennifer Brady, 25, who was in her first tournament. PAGE 2

SUNDAY STYLES
St. Roll Over, Wait Your Turn
Dog trainers are overwhelmed, thanks to a boom in pet adoptions and sales spurred by work-from-home policies and social isolation. PAGE 1

SUNDAY BUSINESS
Boredom Is Making Us Buy
Yet another pandemic side effect: consumers who are eager to make purchases, and not just of groceries. Where sales at Walmart.com. PAGE 1

SUNDAY REVIEW
Katie Engelhart
PAGE 1

Wall of Grief

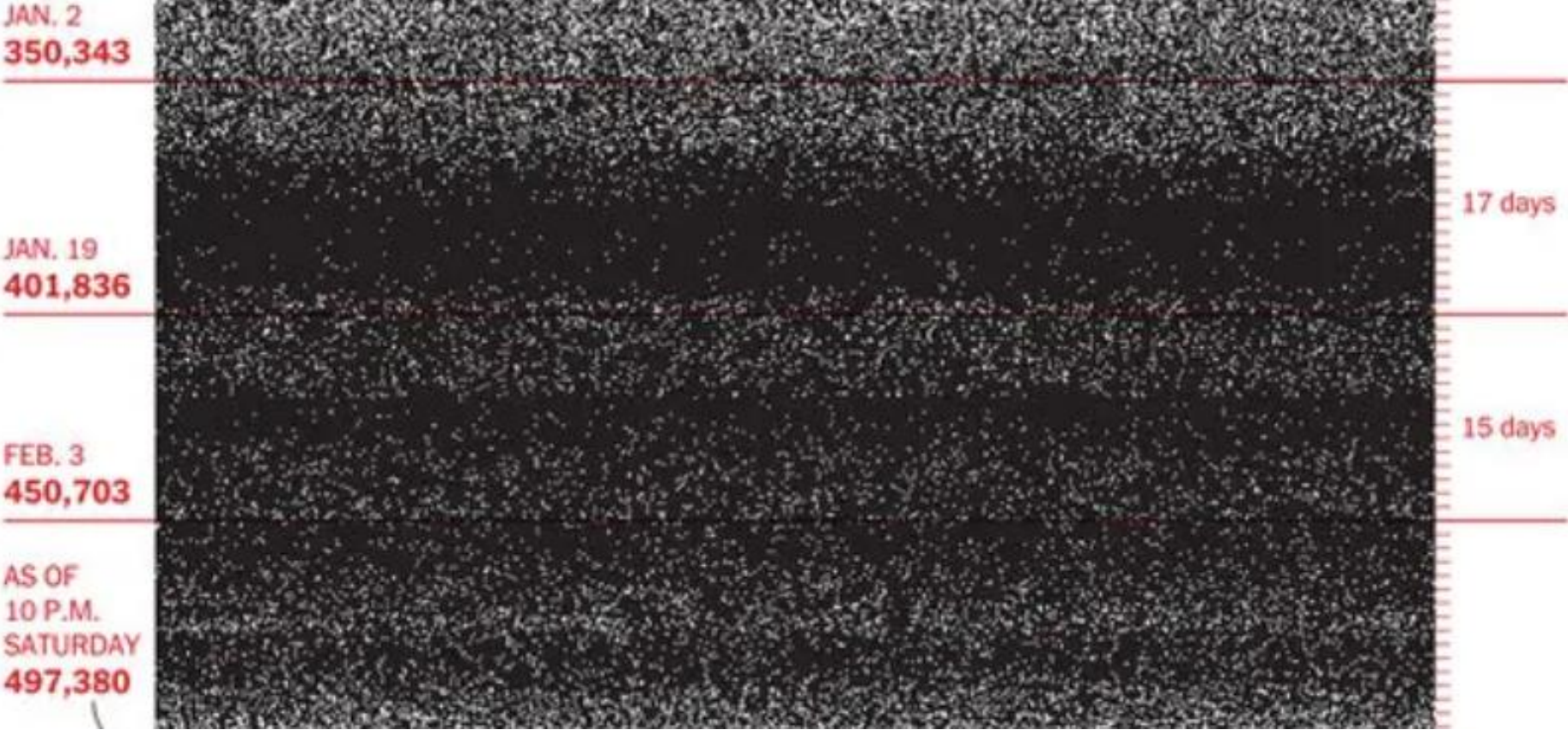
People who have died from COVID-19 in the US

Credit:

The New York Times

<https://www.nytimes.com/2021/02/21/insider/covid-500k-front-page.html>

Zooming in



Where does this visualisation fall?

- Personal → Detached
- Granular → Zoomed out
- Variable → Uniform

What does this visualisation show?

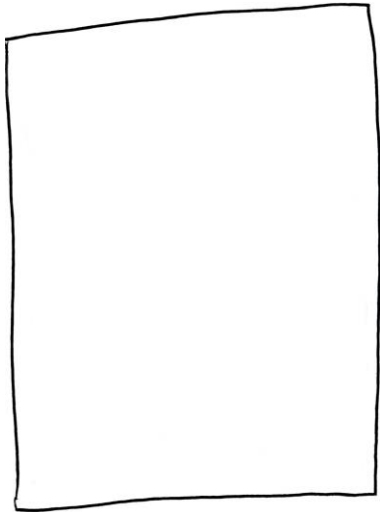
- Enormity
- Monotony
- Speed of change

In summer 2021, I tried to
recreate the wall.

Part 3: My first function

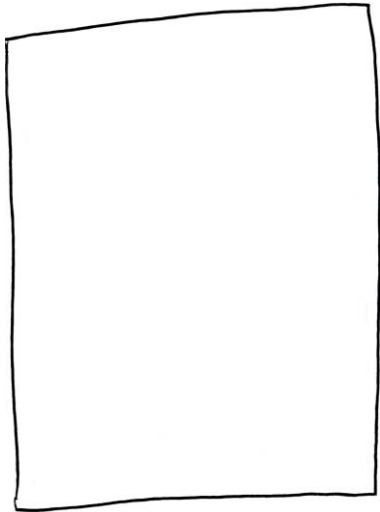
How is this visualisation built?

1

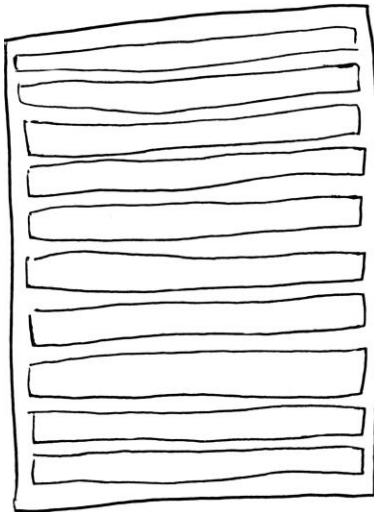


How is this visualisation built?

1

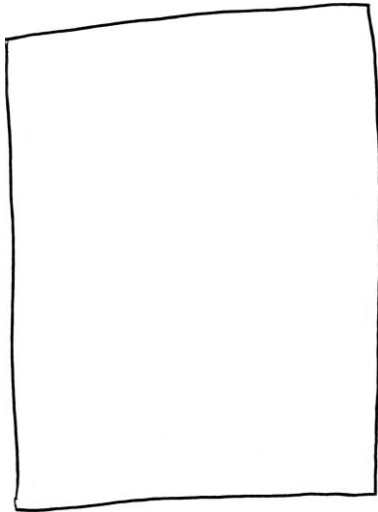


2

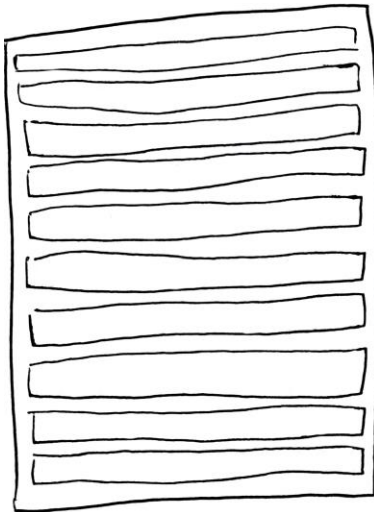


How is this visualisation built?

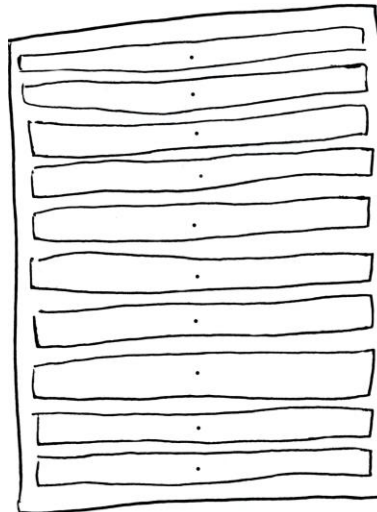
1



2

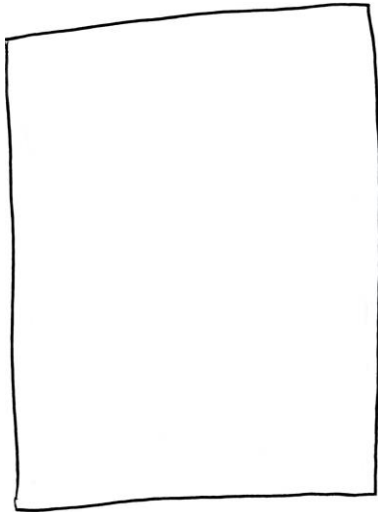


3

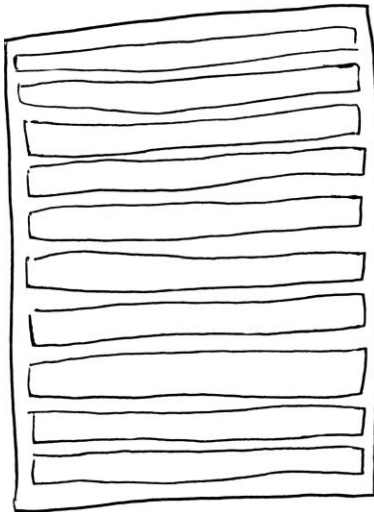


How is this visualisation built?

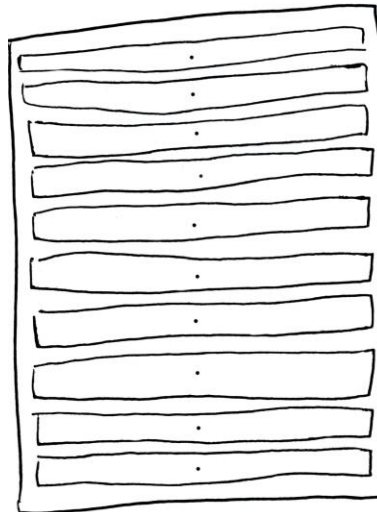
1



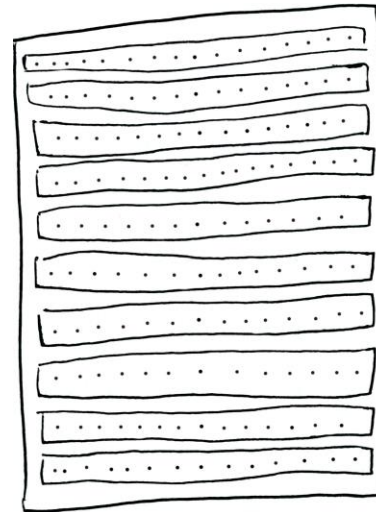
2



3

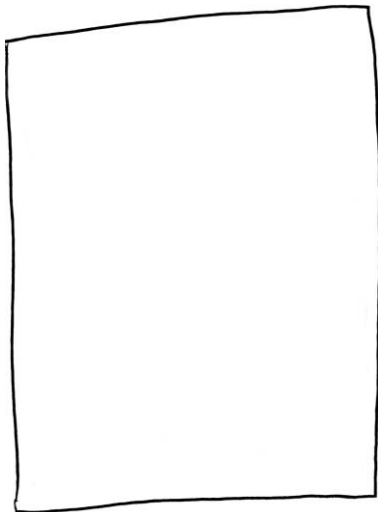


4

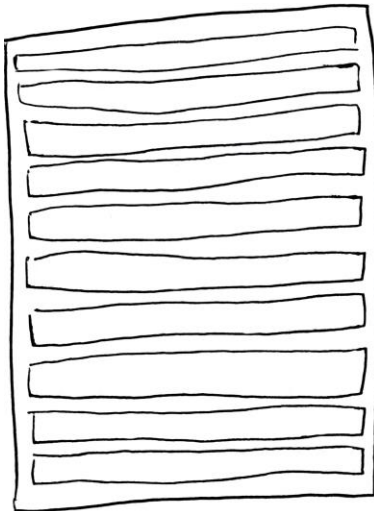


How is this visualisation built?

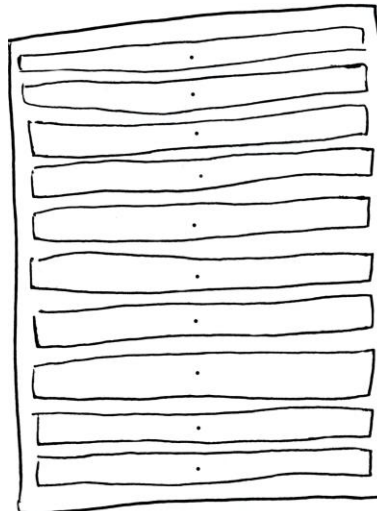
1



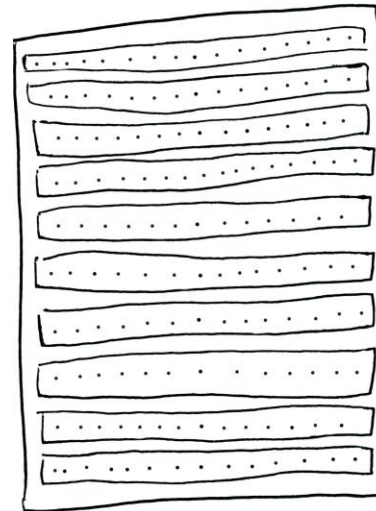
2



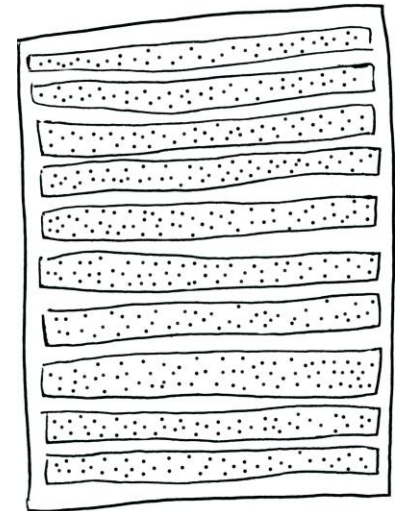
3



4



5



Naming the function

```
onemorelight <- function(df, datetime, title){  
}
```

Loading packages

```
onemorelight <- function(df, datetime, title){  
  library(ggplot2)  
  # functions for customisable data visualisation  
}
```

Converting data to dates

```
onemorelight <- function(df, datetime, title){  
  df[, 1] <- as.POSIXct(strptime(df[, 1], format =  
datetime))  
  # allows users to specify date format (for  
example, decimal month/decimal day/decimal year)  
}
```

Making new data frame for plotting

```
onemorelight <- function(df, datetime, title){  
  duptimes <- df[, 2]  
  index <- rep(1:nrow(df), duptimes)  
  incomp_df <- df[index, ]  
  new_df <- incomp_df  
  # with the number of rows equal to the sum of all  
  events across all days  
}
```

Generating positions for dots

```
onemorelight <- function(df, datetime, title){  
  new_df[, 2] <- runif(nrow(incomp_df), min = 0,  
    max = 10000)  
  new_df[, 3] <- runif(nrow(incomp_df), min =  
0,    max = 100)  
  # using random numbers to prevent overlap  
}
```

Plotting part 1: making dots

```
onemorelight <- function(df, datetime, title){  
  plot <- ggplot(new_df, aes(x = new_df[, 2], y =  
new_df[, 3])) +  
  geom_point(shape = 21, size = 0.02, color =  
"#edd290", fill = "white", stroke = 0.15)  
# very small yellow dots with barely visible  
  outlines  
}
```


Plotting part 2: separating days

```
onemorelight <- function(df, datetime, title){  
  plot <- ggplot(new_df, aes(x = new_df[, 2], y =  
new_df[, 3])) +  
  facet_grid(new_df[, 1] ~.)  
# making one mini-plot for each date  
}
```

Plotting part 3: adding a title

```
onemorelight <- function(df, datetime, title){  
  plot <- ggplot(new_df, aes(x = new_df[, 2], y =  
new_df[, 3])) +  
  labs(title = title)  
# title specified by user  
}
```

Plotting part 4: adding an overall theme

```
onemorelight <- function(df, datetime, title){  
  plot <- ggplot(new_df, aes(x = new_df[, 2], y =  
new_df[, 3])) +  
  theme_void()  
# to divorce the visual from a typical graph  
}
```

Plotting part 5: additional customisation

```
onemorelight <- function(df, datetime, title){  
  plot <- ggplot(new_df, aes(x = new_df[, 2], y =  
new_df[, 3])) +  
  theme(plot.title = element_text(face = "bold", color =  
"white", size = 5, margin = margin(0, 0, 5, 0), hjust =  
0.055), strip.text = element_text(size = 2.5, color =  
"light grey", face = "bold", margin = margin(2, 0, 3,  
0)))  
# customising title and y-axis labels  
}
```

(cont.)

```
onemorelight <- function(df, datetime, title){  
  plot <- ggplot(new_df, aes(x = new_df[, 2], y =  
new_df[, 3])) +  
  theme(plot.margin = margin(20, 20, 20, 20, "pt"),  
        plot.background = element_rect(fill = "#002241",  
color = NA),  
        panel.spacing = unit(0.008, "lines"))))  
# customising margins, background, and spacing  
}
```

Plotting part 6: displaying the plot

```
onemorelight <- function(df, datetime, title){  
  return(plot)  
# displays in R plot viewer  
}
```

Can we do better?

- Conciseness
- Clarity
- Flexibility

Part 4: Making a better function

```
onemorelight <- function(df, ...){  
}
```


Running data checks: first column

```
onemorelight <- function(df, ...){  
  if (class(df[, 1]) != "Date" && class(df[, 1]) !=  
"numeric") {  
    abort(paste0(  
      "First column must be a Date or numeric, not ",  
      typeof(df[, 1]), ", object."  
    )) # more flexible to Date or numeric classes  
    # stops execution if data types are incorrect  
  }  
}
```

Running data checks: second column

```
onemorelight <- function(df, ...){  
  if (class(df[, 2]) != "numeric") {  
    abort(paste0(  
      "First column must be a numeric, not ",  
      typeof(df[, 2]), ", object."  
    ))  
    # stops execution if data types are incorrect  
  }  
}
```

Making new data frame for plotting

```
onemorelight <- function(df, ...){  
  plot_df <- data.frame(date = df[rep(1:nrow(df),  
                                df[, 2]), 1], # number of  
                        total days x number of events on each day  
                        x = runif(sum(df[, 2]), min = 0, max  
= 10000),  
                        y = runif(sum(df[, 2]), min = 0, max  
= 100)  
                        )  
}
```

Plotting part 1: making dots

```
onemorelight <- function(df, dot_colour, dot_fill){  
  ggplot(plot_df, aes(x = x, y = y)) +  
    geom_point(shape = 21, size = 0.02, colour =  
dot_colour, fill = dot_fill, stroke = 0.15)  
# identical structure as previous with  
additional colour options  
}
```

Plotting part 2: separating days

```
onemorelight <- function(df, facet){  
  ggplot(plot_df, aes(x = x, y = y)) +  
    {  
      if (facet) facet_grid(rows = vars(date))  
    }  
  # separates days depending on user specification  
}
```

Plotting part 3: adding a title

```
onemorelight <- function(df, title){  
  ggplot(plot_df, aes(x = x, y = y)) +  
  labs(title = title)  
# identical to previous  
}
```

Plotting part 4: adding an overall theme

```
onemorelight <- function(df){  
  ggplot(plot_df, aes(x = x, y = y)) +  
  theme_void()  
# identical to previous  
}
```

Plotting part 5: additional customisation

```
onemorelight <- function(df, title_colour, label_colour){  
  ggplot(plot_df, aes(x = x, y = y)) +  
    theme(plot.title = element_text(face = "bold", color =  
title_colour, size = 5, margin = margin(0, 0, 5, 0),  
hjust = 0.055), strip.text = element_text(size = 2.5,  
color = label_colour, face = "bold", margin = margin(2,  
0, 3, 0)))  
# identical structure as previous with additional colour  
  customisation  
}
```


(cont.)

```
onemorelight <- function(df, bg_fill){  
  ggplot(plot_df, aes(x = x, y = y)) +  
  theme(plot.margin = margin(20, 20, 20, 20, "pt"),  
        plot.background = element_rect(fill = bg_fill,  
color = NA),  
        panel.spacing = unit(0.008, "lines"))))  
# identical structure as previous with additional  
  colour customisation  
}
```

Plotting part 6: exporting plot

```
onemorelight <- function(df, filename, width,  
  height){  
  ggsave(filename, width = width, height = height)  
  # wide range of file formats available  
}
```

Function arguments and defaults

```
(df, # data frame with dates and events per date
dot_colour = "#edd290", # colour of dot outlines
dot_fill = "white", # colour of dot fills
bg_fill = "#002241", # colour of plot background
title_colour = "white", # colour of title text
label_colour = "white", # colour of label text
facet = T, # whether to separate events by date
title = "one more light", # plot title
filename, # file name of exported plot
width = 30, height = 45) # width and height of exported plot
```

Things to remember

- Don't use confusing names (`new_df`, `first_df`)
- Generate clear error messages (that didn't work, here's why)
- Specify data and package requirements beforehand
- Condense code as far as possible
- Use defaults where useful
- Make outputs customisable
- Enable users to save outputs

Part 5: An example

```
onemorelight(df)
```

Loading our packages

```
library(ggplot2)
# highly customisable functions for plotting
library(rlang)
# functions for working with R core language
features
```

Loading our data

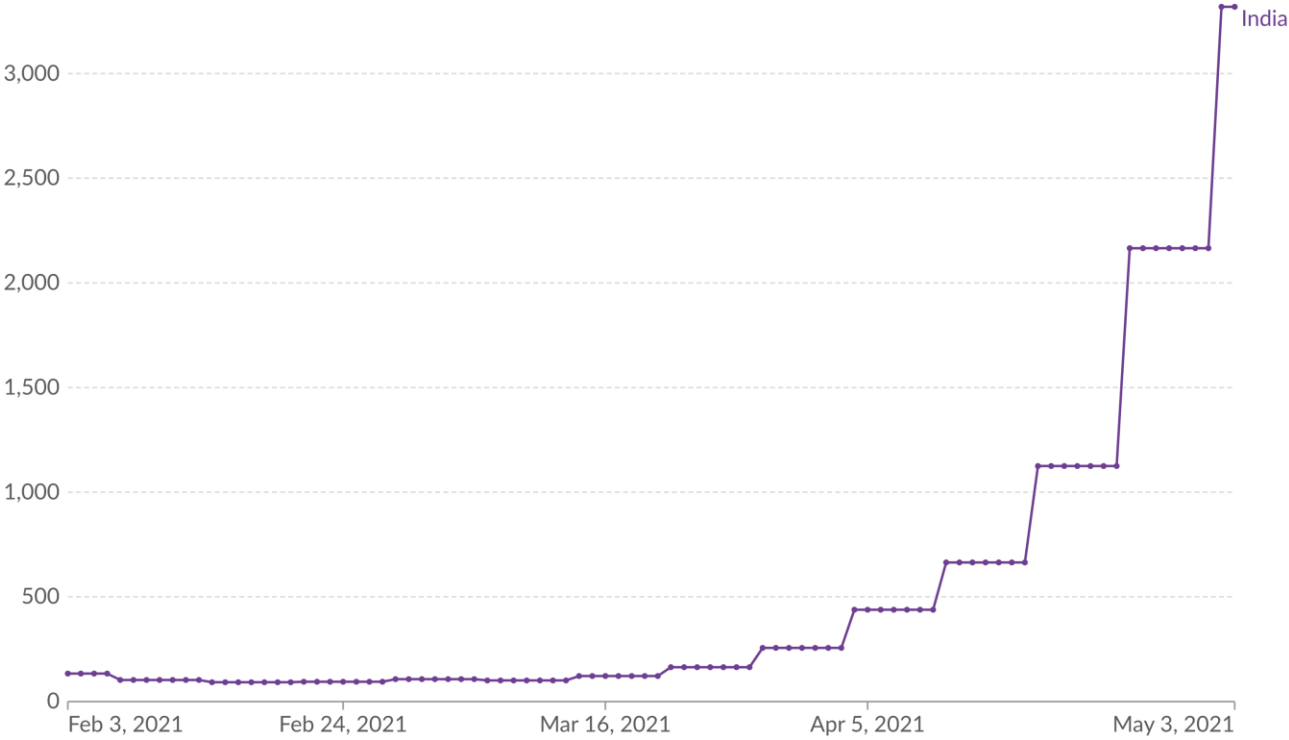
```
df <- read.csv(onemorelight_data.csv")  
# data from the WHO COVID-19 Dashboard showing daily  
deaths from COVID-19 in India in Feb-May 2021
```

A conventional representation of this data

Daily new confirmed COVID-19 deaths



7-day rolling average. Due to varying protocols and challenges in the attribution of the cause of death, the number of confirmed deaths may not accurately represent the true number of deaths caused by COVID-19.



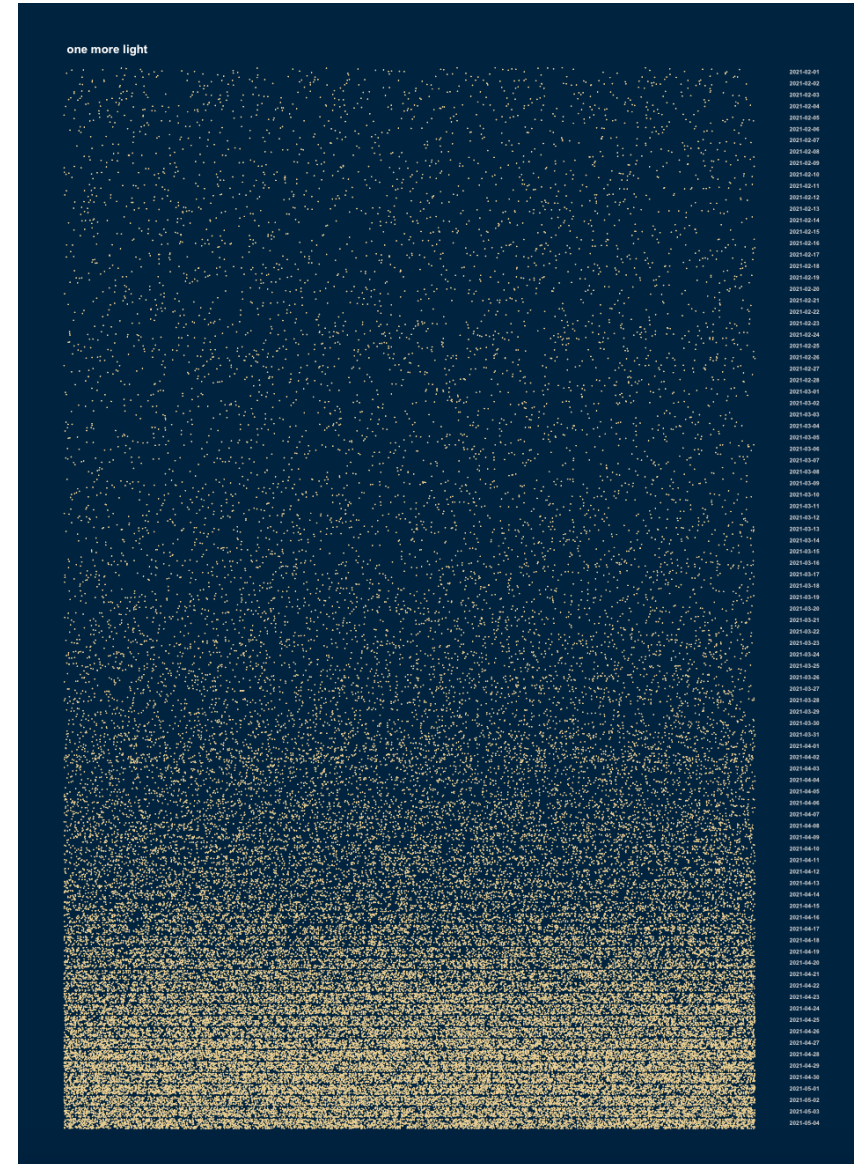
Data source: WHO COVID-19 Dashboard

[CC BY](#)

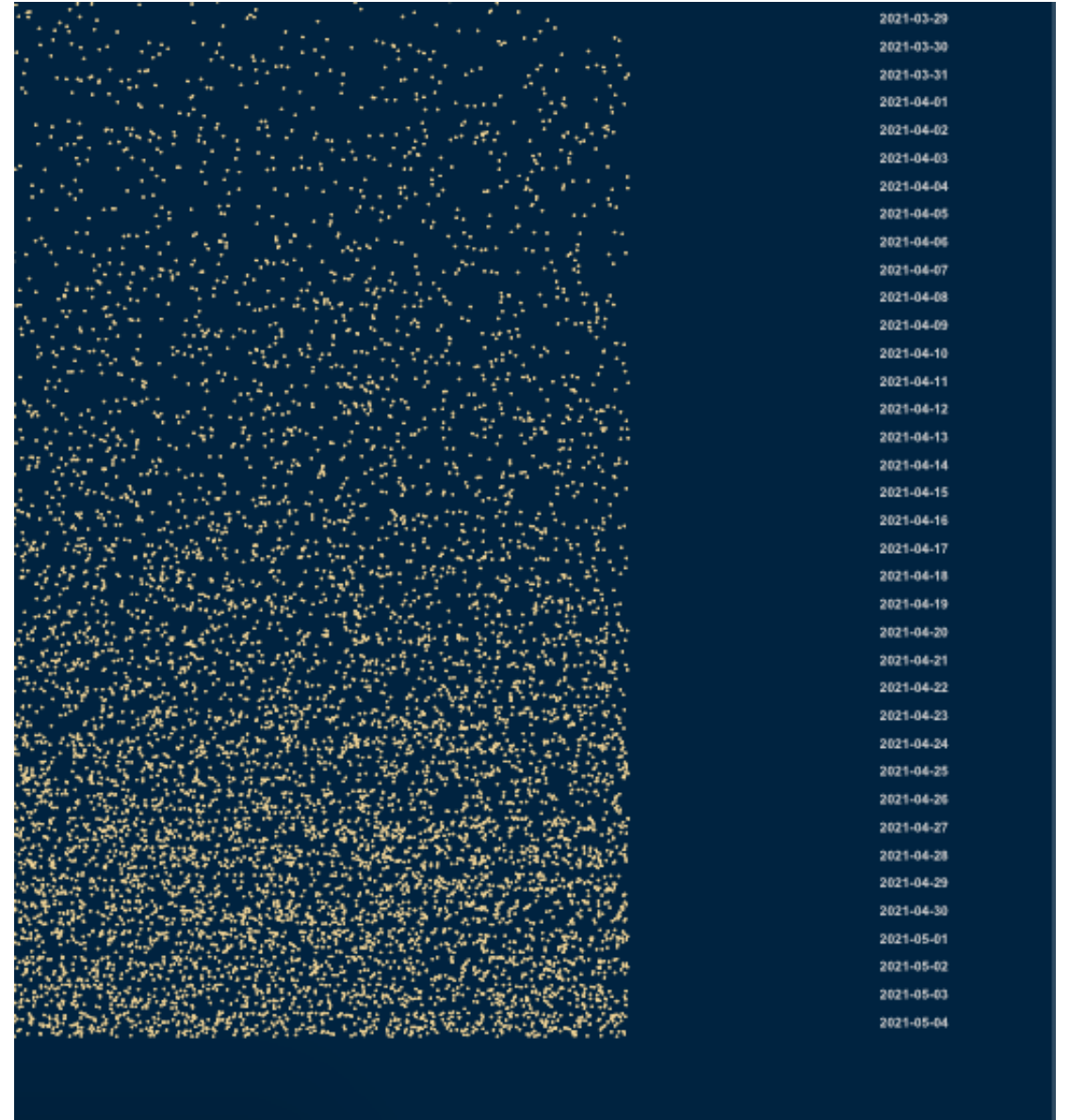
How can we show the data differently?

```
onemorelight(df, "onemorelight.png")  
# using design defaults
```

Final product



Zooming in



Part 6: Reflecting on the process

Can R play a role?

- What can R create?
- What do we usually ask of code? How important are those asks?



Cherish life everyday

People alive today because of a volunteer who patrols a Chinese river bridge

Credit:

@wmjluh (Medium)

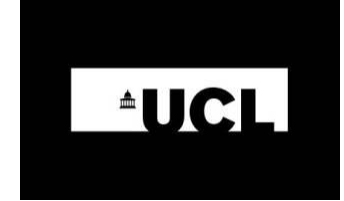
Thank you

Acknowledgements

one more light



Academic colleagues



Professional colleagues



The Data Visualization Society



Friends and family



Keep in touch

Yaning Wu (she/her)

yw645@medschl.cam.ac.uk

PhD student in blood donor health
Blood and Transplant Research Unit
University of Cambridge

 yaning-wu

 nerdfighterwu

 yaning-wu

 wmjlwuh

