



#30DayChartChallenge

Nicola Rennie

26 May 2022

About me

- PhD Statistics and Operational Research
- Data scientist at Jumping Rivers
 - Consultancy: shiny, statistics, slides, ...
 - Internal projects: blogdown websites, reports, plot styling, admin, ...
 - Training: all things R (and some Tableau coming soon...)
- A lot of data visualisation...



Nicola Rennie

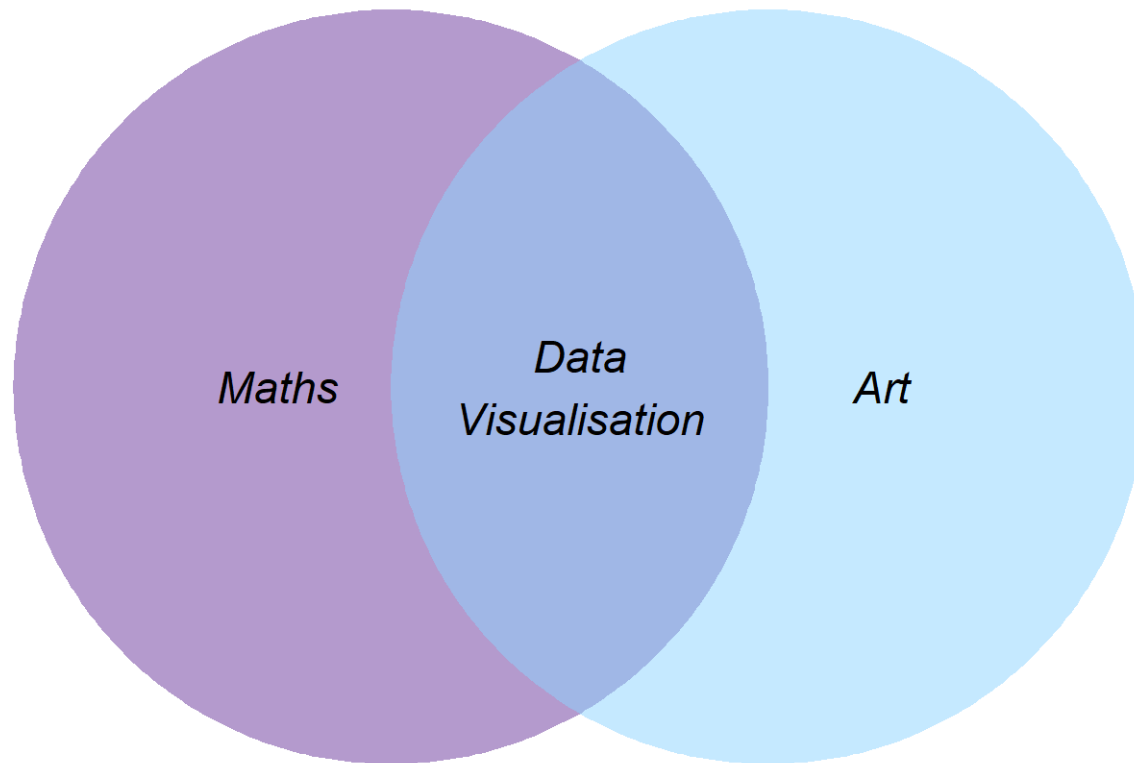
My R Journey

- Compulsory R coursework for a statistics course during undergraduate degree
- Learnt Python instead...
- Final year of undergrad gave R another go
- Started learning {tidyverse} during PhD



Nicola Rennie

Why data visualisation?



Part 1

#30DayChartChallenge

What is the #30DayChartChallenge?






- Data visualisation challenge where participants make one chart each day inspired by a daily prompt and category.
- Organised by [Cédric Scherer](#) and [Dominic Royé](#), with support from [Wendy Shijia](#) and [Marco Sciaini](#).
- Post charts on Twitter with the #30DayChartChallenge (and #DayX)
- See also: 30daychartchallenge.org




Prompts

#30DayChartChallenge

April 2022 • 30 Days • 30 Charts • 5 Categories



Comparisons	Distributions	Relationships	Timeseries	Uncertainties
1. part-to-whole	7. physical	13. correlation	19. global change	25. trend
2. pictogram	8. mountains	14. 3-dimensional	20. new tool	26. interactive
3. historical	9. statistics	15. multivariate	21. down/upwards	27. future
4. flora	10. experimental	16. environment	22. animation	28. deviations
5. slope	11. circular	17. connections	23. tiles	29. storytelling
6. data day: OWID	12. theme day: The Economist	18. data day: OECD	24. theme day: Financial Times	30. data day: UN Population

Follow  @30DayChartChall for more!



Nicola Rennie

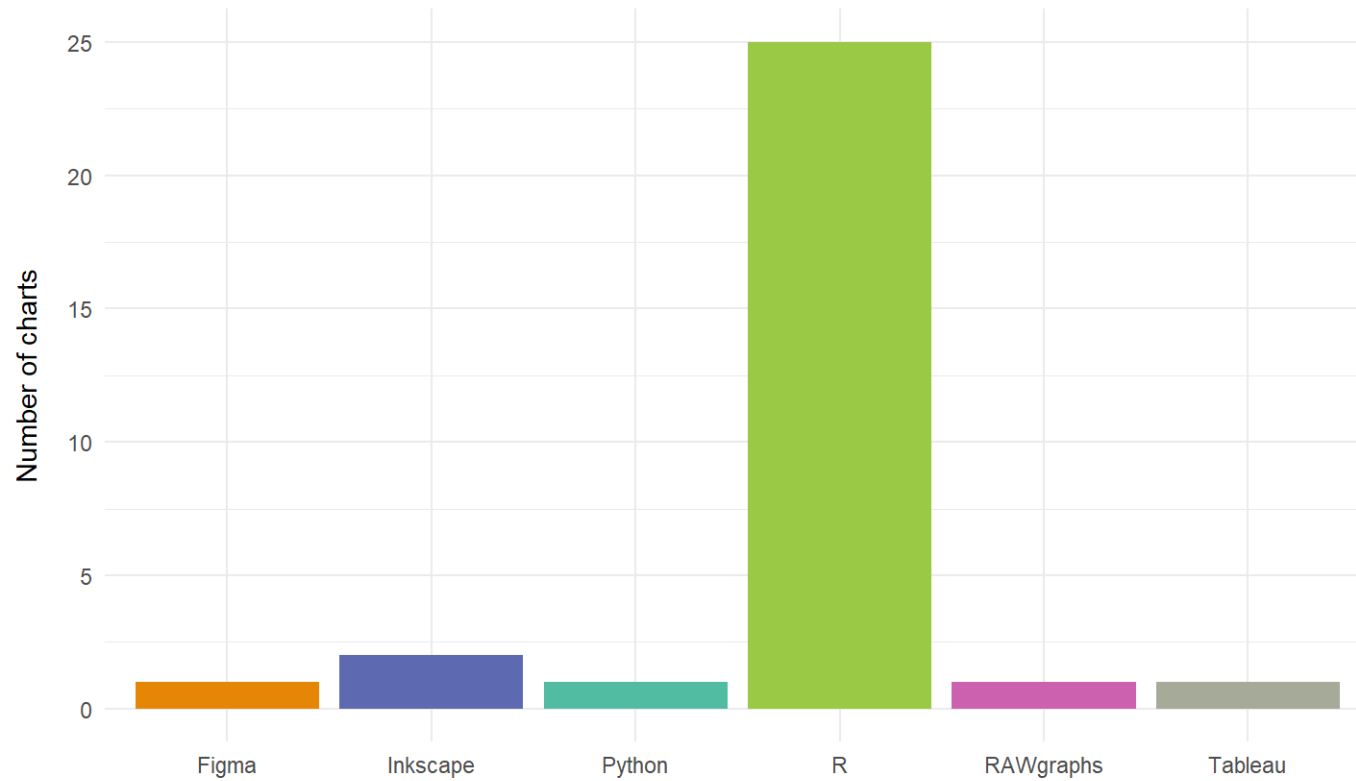
Why did I make 30 charts?

- One “new tool” for each of the five categories
- Learn some new things
- Make charts that I wanted to make
- Have fun!



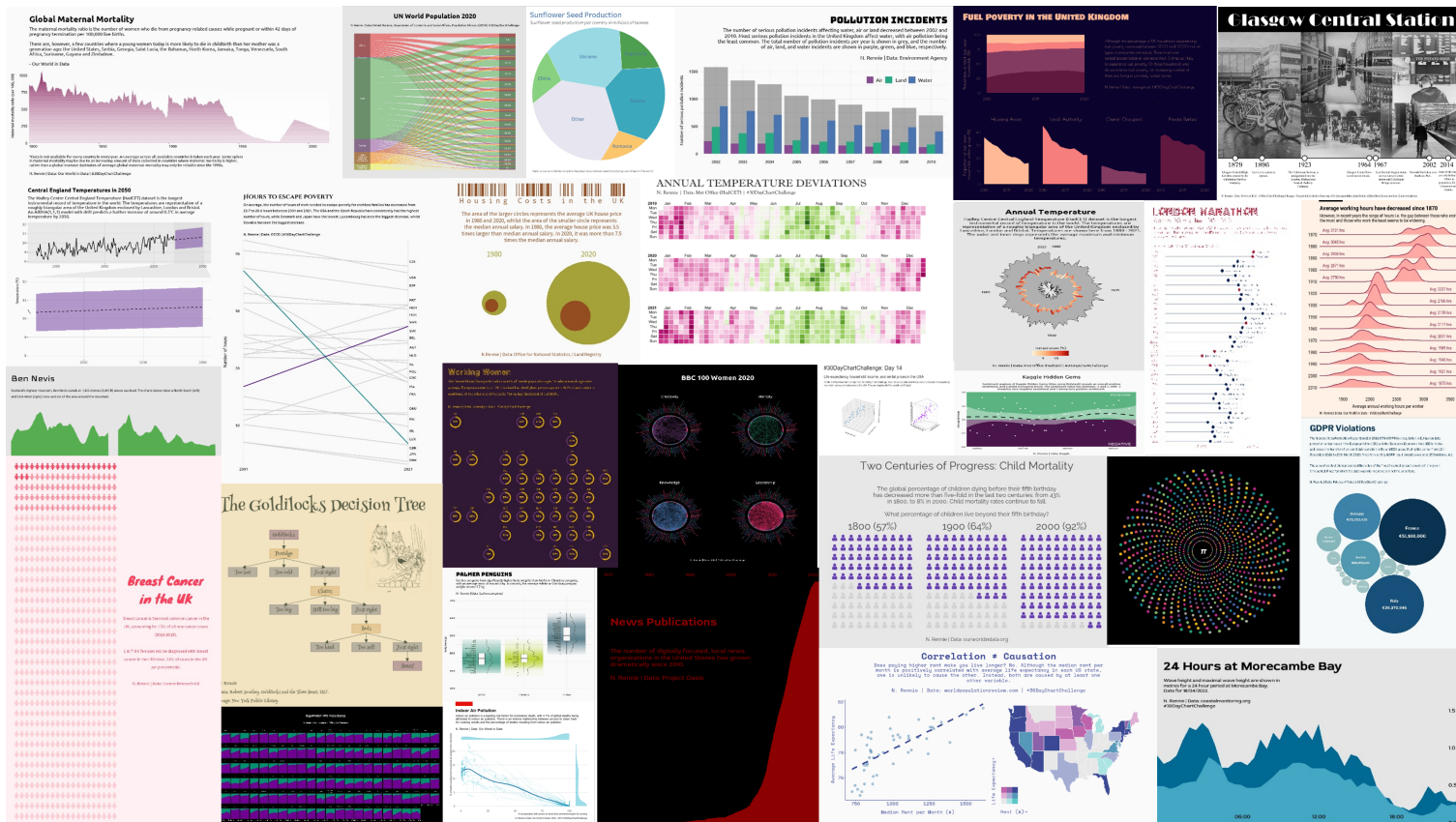
Nicola Rennie

How did I make 30 Charts?



Nicola Rennie

The 30 Charts



Nicola Rennie

Day 1 (Part to whole) in R

Working Women

The female labour force participation rate (% of female population ages 15+ who are working) varies amongst European countries. In 2019, Iceland had the highest percentage with 70.2% of women in the workforce. At the other end of the scale, Turkey had the lowest at just 34.3%.

N. Rennie | Data: Diversity in Data | #30DayChartChallenge



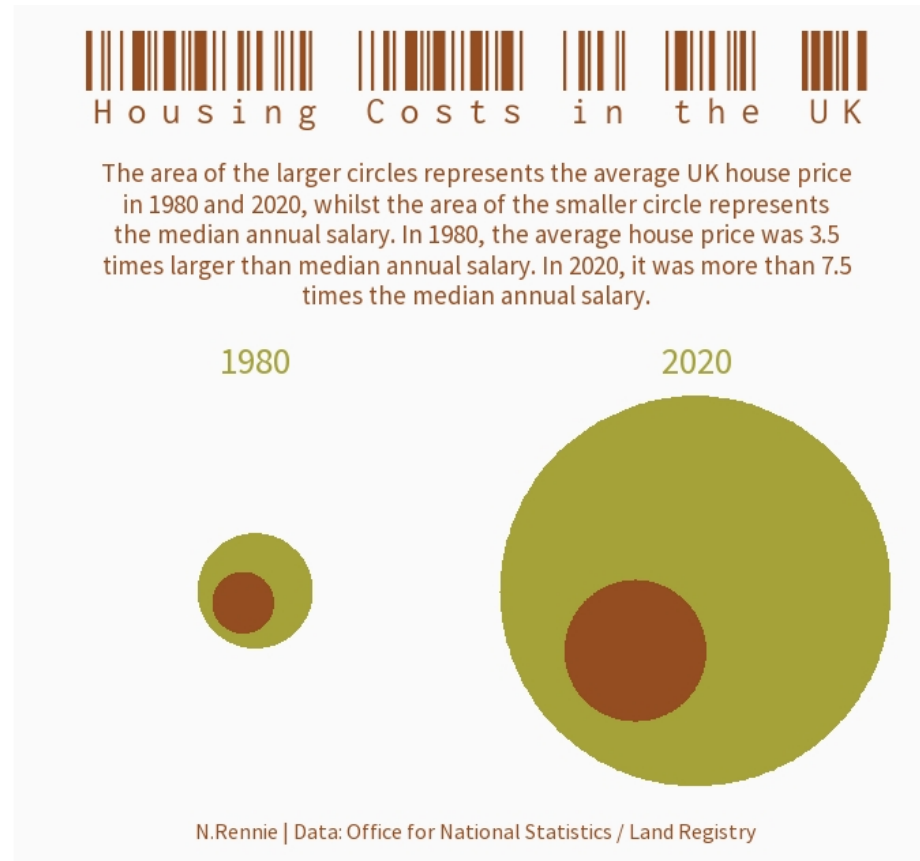
Nicola Rennie

Day 2 (Pictogram) in R



Nicola Rennie

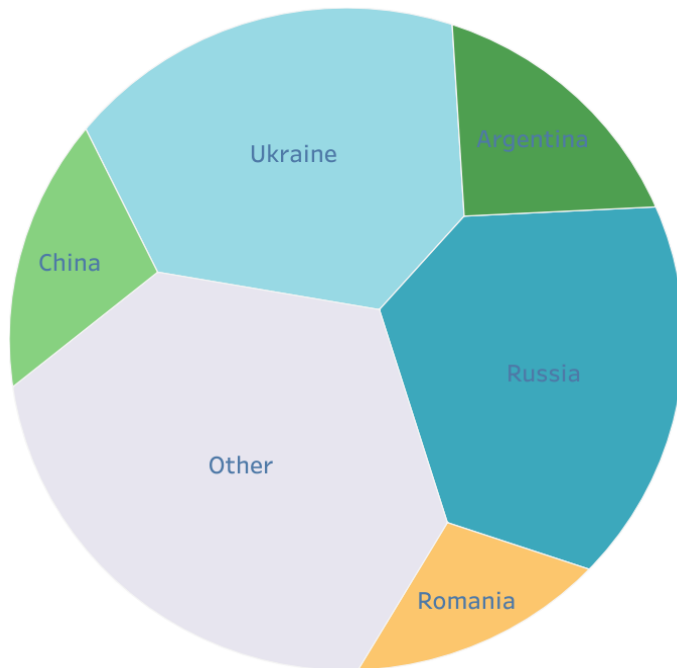
Day 3 (Historical) in R



Day 4 (Flora) in Tableau (left) and R (right)

Sunflower Seed Production

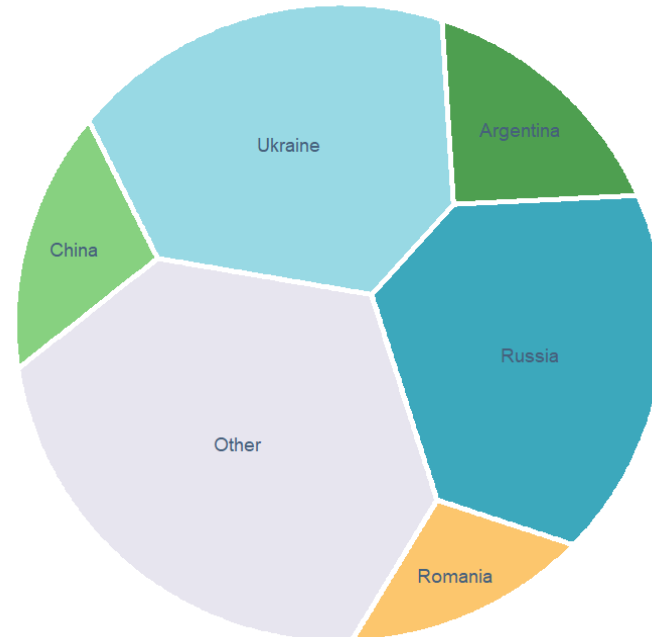
Sunflower seed production per country in millions of tonnes.



Data: www.worldatlas.com/articles/the-top-sunflower-seed-producing-countries-in-the-world

Sunflower Seed Production

Sunflower seed production per country in millions of tonnes.



Data: www.worldatlas.com/articles/the-top-sunflower-seed-producing-countries-in-the-world



Day 5 (Slope) in R



Nicola Rennie

Day 6 (Our World in Data) in R

Two Centuries of Progress: Child Mortality

The global percentage of children dying before their fifth birthday has decreased more than five-fold in the last two centuries: from 43% in 1800, to 8% in 2000. Child mortality rates continue to fall.

What percentage of children live beyond their fifth birthday?



N. Rennie | Data: ourworldindata.org



Nicola Rennie

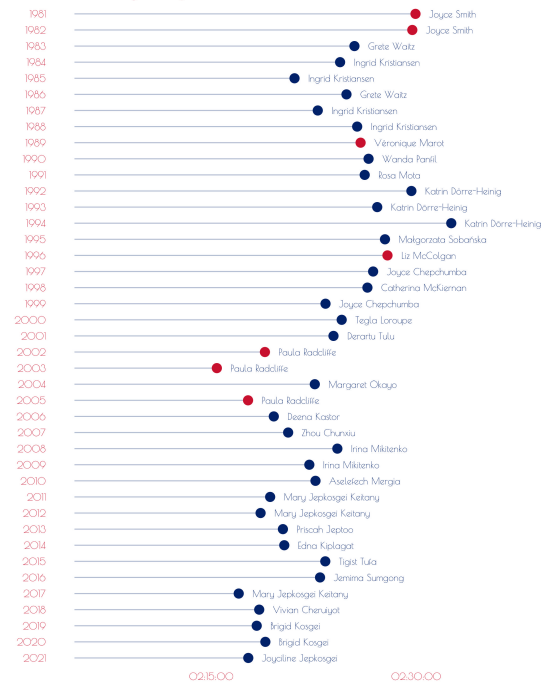
Day 7 (Physical) in R

LONDON MARATHON

Womens' Winning Times 1981 - 2021

The London Marathon began in 1981, and 27 different women have won. The race has been won by a British runner 7 times. In 2003, Paula Radcliffe won in a time of 2:15:25, a world record which would stand for 16 years.

N. Rennie | Data: Virgin Money London Marathon



Nicola Rennie

Day 8 (Mountains) in Figma

Ben Nevis

Scotland's highest mountain, Ben Nevis stands at 1,345 metres (4,413 ft) above sea level. The charts below show a North-South (left) and East-West (right) cross-section of the area around the mountain.

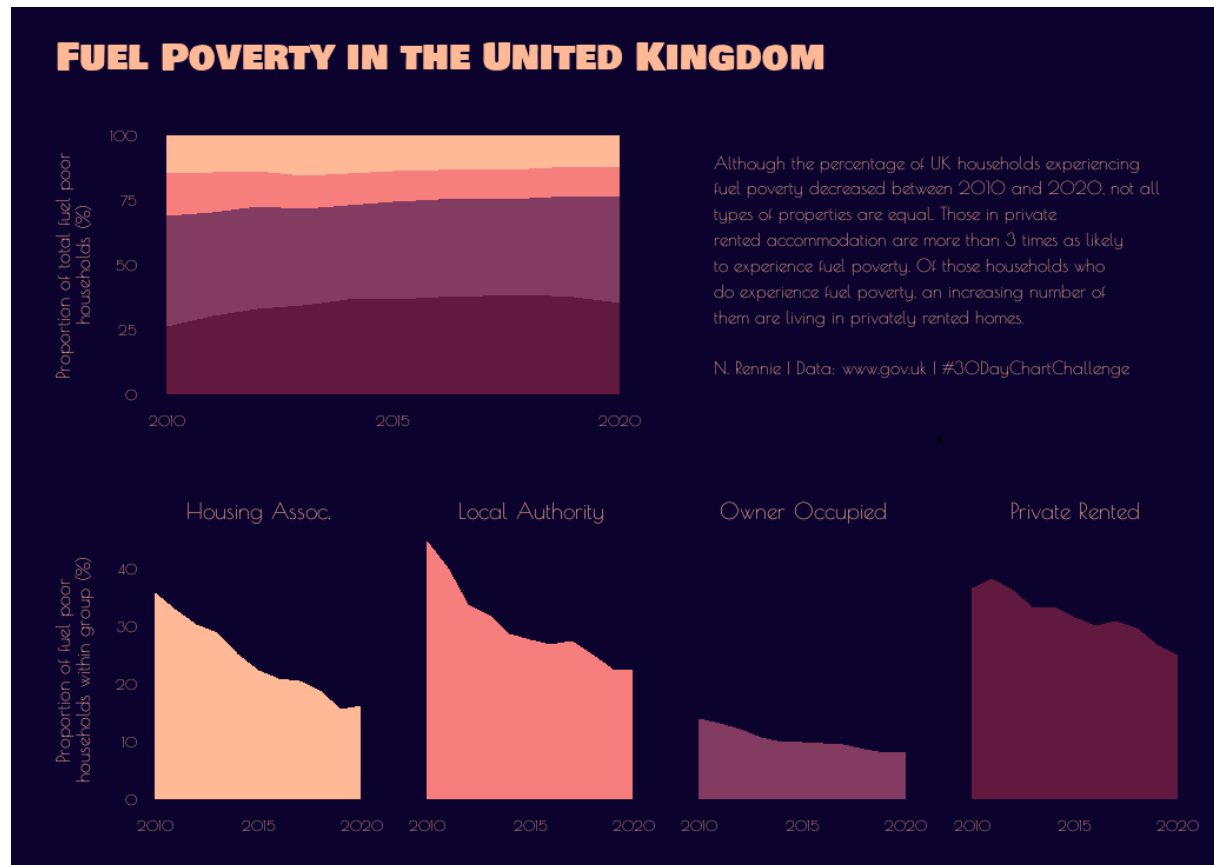


N. Rennie | Data: OS Datahub | #30DayChartChallenge



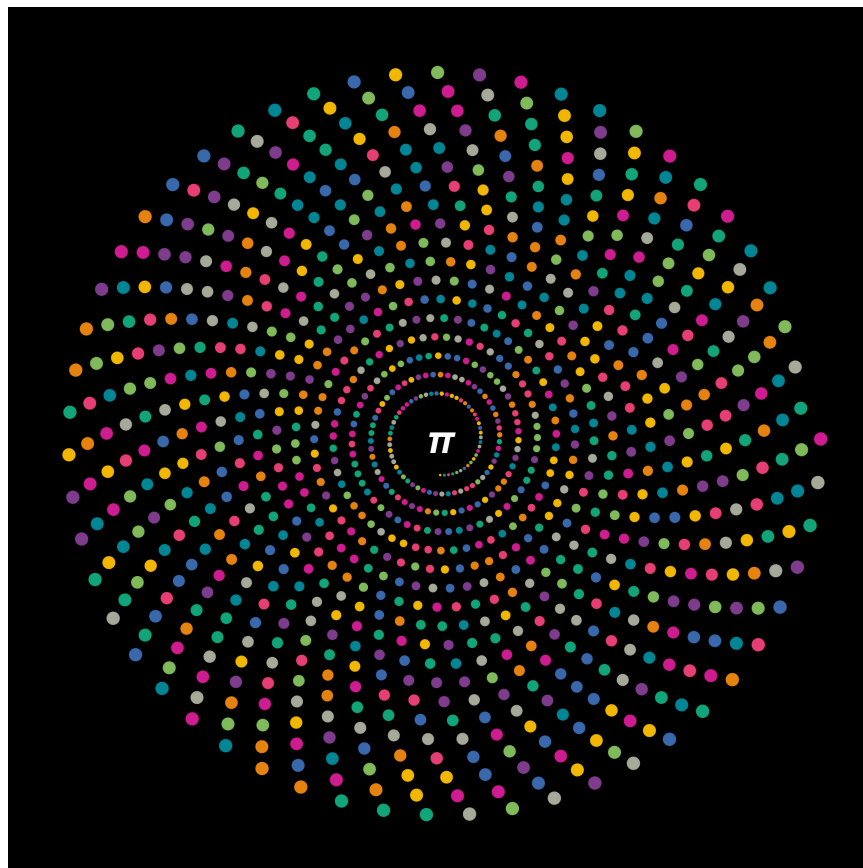
Nicola Rennie

Day 9 (Statistics) in R



Nicola Rennie

Day 10 (Experimental) in R



Nicola Rennie

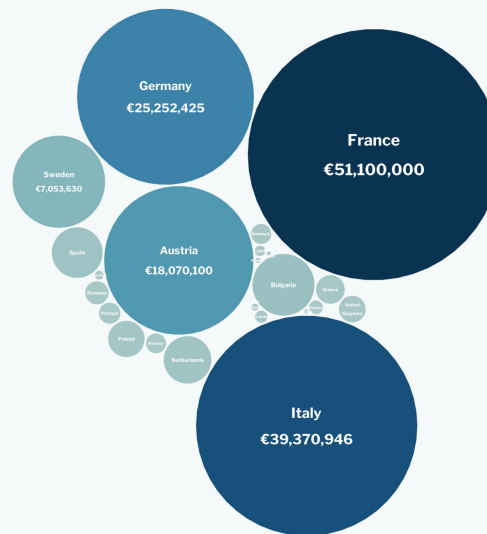
Day 11 (Circular) in R

GDPR Violations

The General Data Protection Regulation (EU) 2016/679 (GDPR) is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA). It also addresses the transfer of personal data outside the EU and EEA areas. During the period from 12th December 2018 to 25th March 2020, fines for violating GDPR requirements exceeded 153 million euros.

The area of each circle represented the value of the fines incurred by each country in the given time period. Fines for which the date was mis-recorded are not included here.

N. Rennie | Data: Privacy Affairs | #30DayChartChallenge



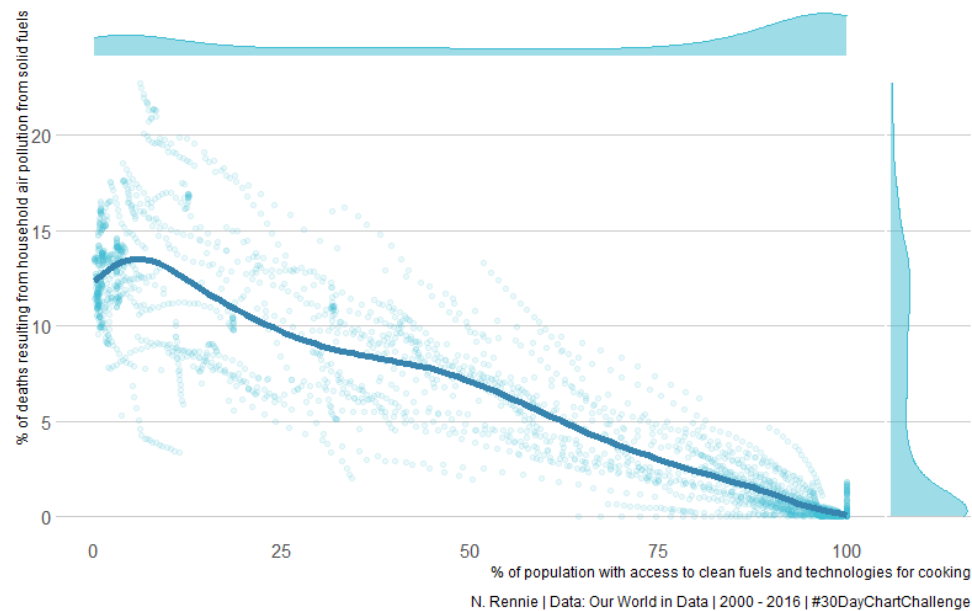
Nicola Rennie

Day 12 (The Economist) in R

Indoor Air Pollution

Indoor air pollution is a leading risk factor for premature death, with 4.1% of global deaths being attributed to indoor air pollution. There is an inverse relationship between access to clean fuels for cooking results and the percentage of deaths resulting from indoor air pollution.

N. Rennie | Data: Our World in Data



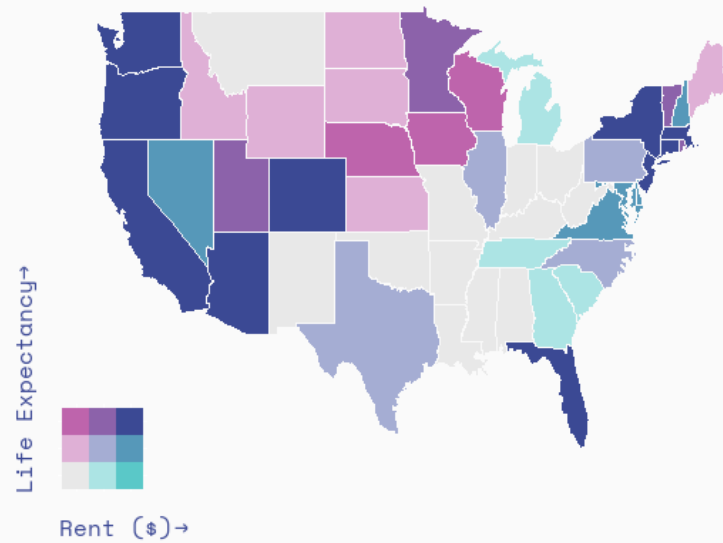
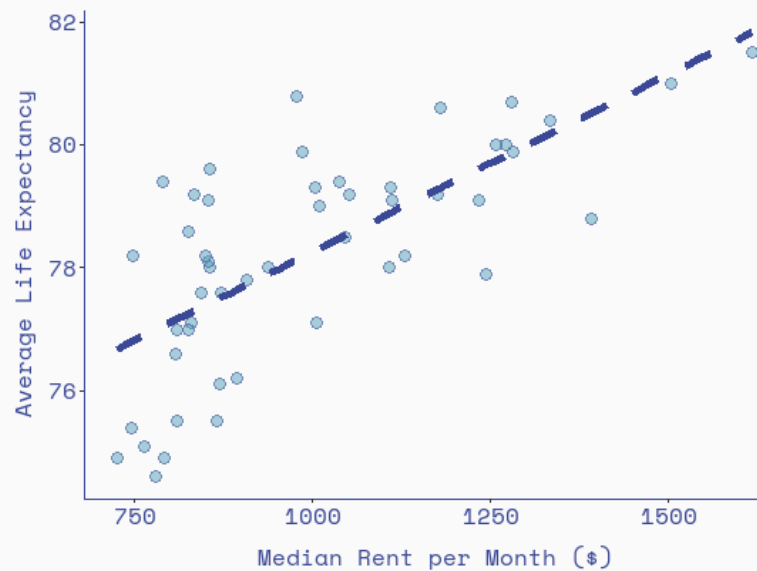
Nicola Rennie

Day 13 (Correlation) in R

Correlation ≠ Causation

Does paying higher rent make you live longer? No. Although the median rent per month is positively correlated with average life expectancy in each US state, one is unlikely to cause the other. Instead, both are caused by at least one other variable.

N. Rennie | Data: worldpopulationreview.com | #30DayChartChallenge



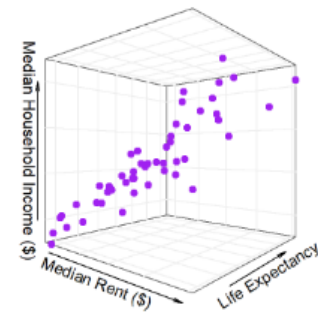
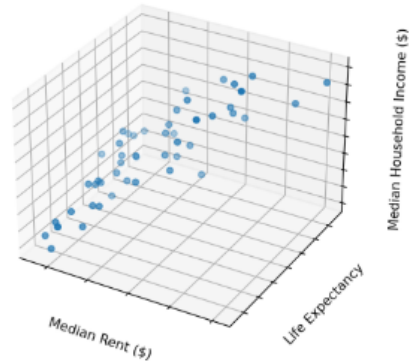
Nicola Rennie

Day 14 (3-Dimensional) in Python and R

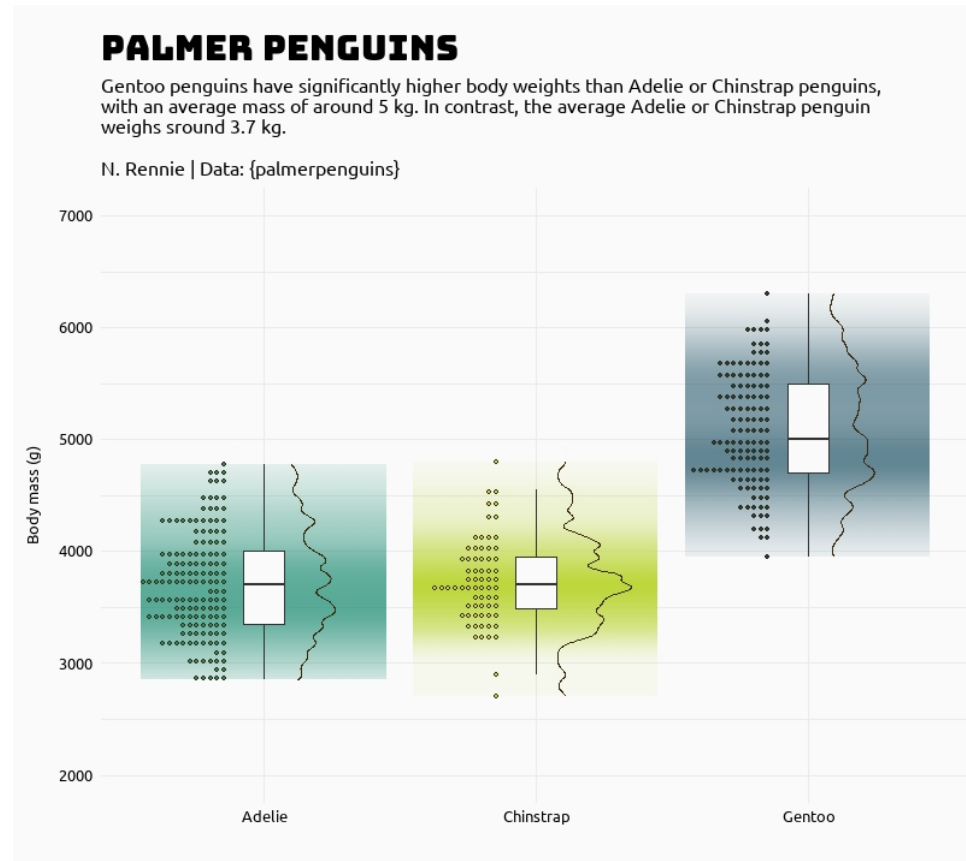
#30DayChartChallenge: Day 14

Life expectancy, household income, and rental prices in the USA

For the *3-dimensional* theme of day 14 of the 30 Day Chart Challenge, these 3D scatter plots show the relationship between life expectancy, household income, and rental prices in the USA. They are visualised in Python (left) and R (right).



Day 15 (Multivariate) in R

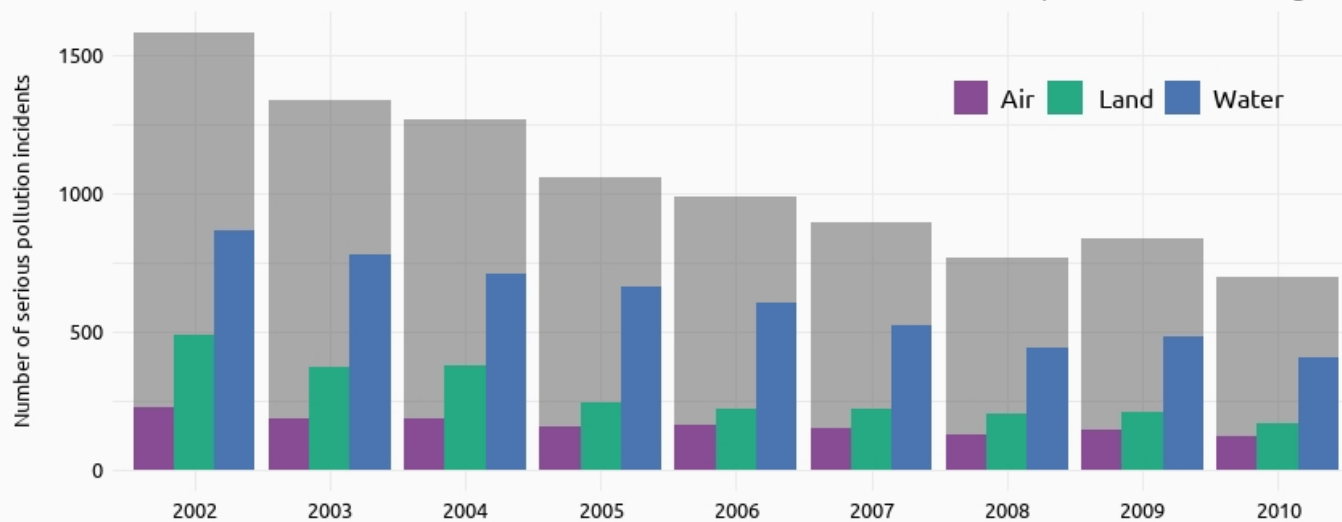


Day 16 (Environment) in R

POLLUTION INCIDENTS

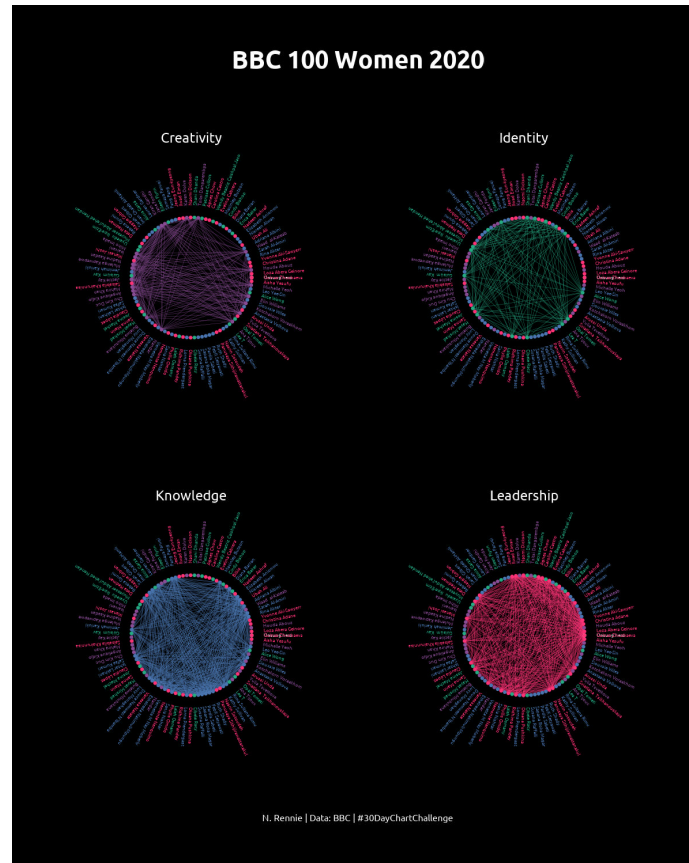
The number of serious pollution incidents affecting water, air or land decreased between 2002 and 2010. Most serious pollution incidents in the United Kingdom affect water, with air pollution being the least common. The total number of pollution incidents per year is shown in grey, and the number of air, land, and water incidents are shown in purple, green, and blue, respectively.

N. Rennie | Data: Environment Agency



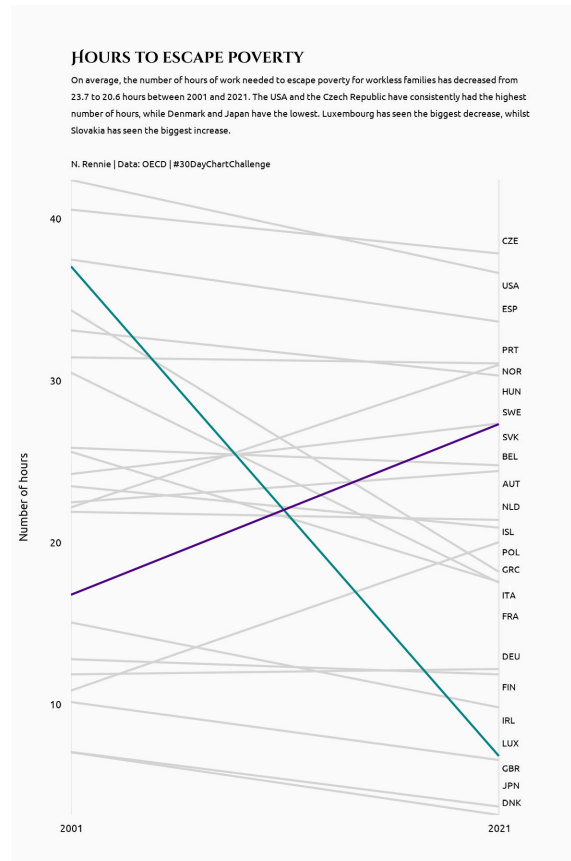
Nicola Rennie

Day 17 (Connections) in R



Nicola Rennie

Day 18 (OECD) in R



Nicola Rennie

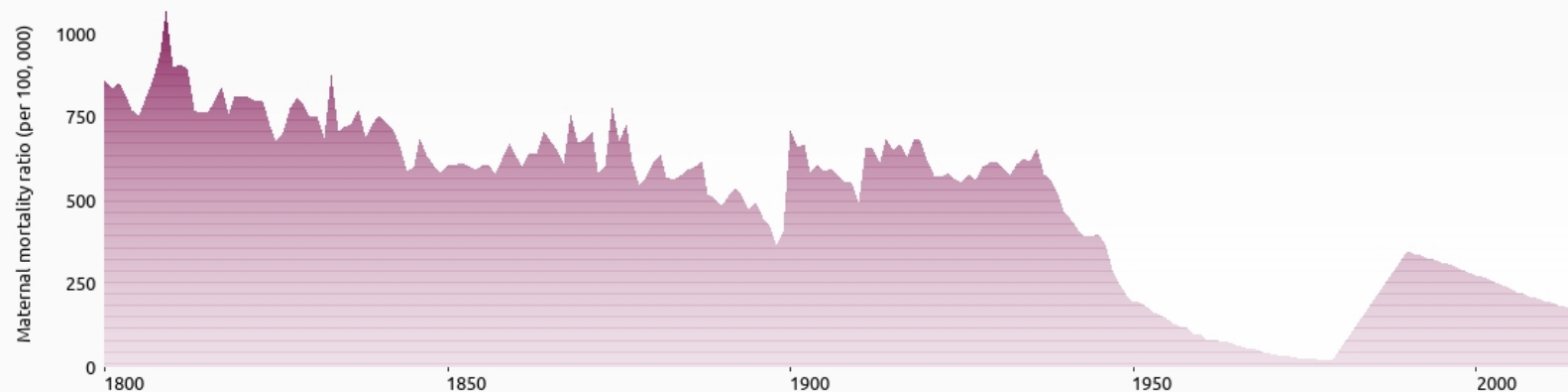
Day 19 (Global Change) in R

Global Maternal Mortality

The maternal mortality ratio is the number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births.

There are, however, a few countries where a young women today is more likely to die in childbirth than her mother was a generation ago: the United States, Serbia, Georgia, Saint Lucia, the Bahamas, North Korea, Jamaica, Tonga, Venezuela, South Africa, Suriname, Guyana and Zimbabwe.

- Our World in Data



*Data is not available for every country in every year. An average across all available countries is taken each year. Some spikes in maternal mortality may be due to an increasing amount of data collected in countries where maternal mortality is higher, rather than a global increase. Estimates of average global maternal mortality may only be reliable since the 1990s.

N. Rennie | Data: Our World in Data | #30DayChartChallenge



Nicola Rennie

Day 20 (New Tool) in Inkscape

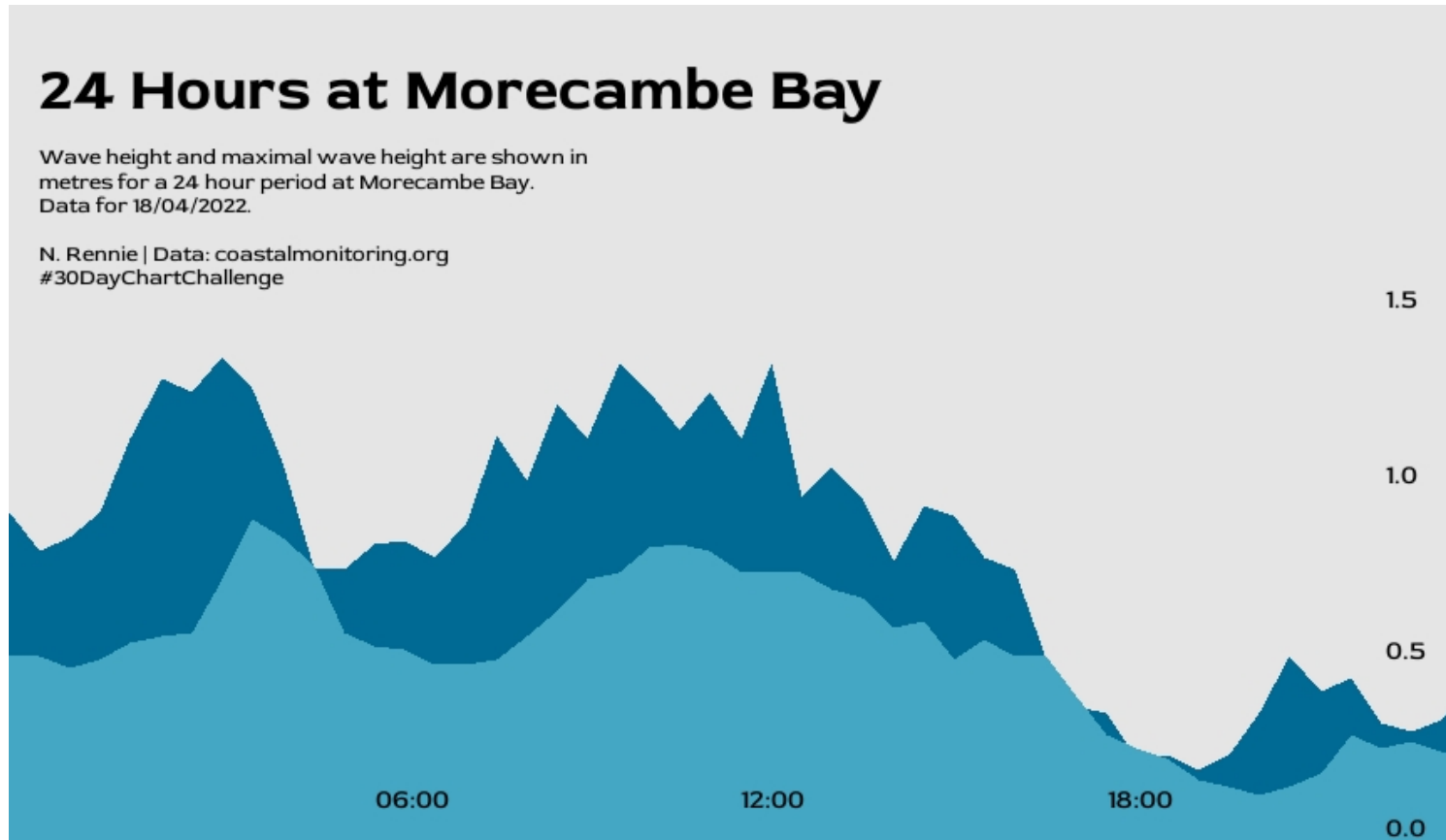


N. Rennie | Data: Network Rail | #30DayChartChallenge | Images: Unsplash (clockwise from top left) @rossneddon @profwicks @ilyailford @rossneddon @andrewjephson



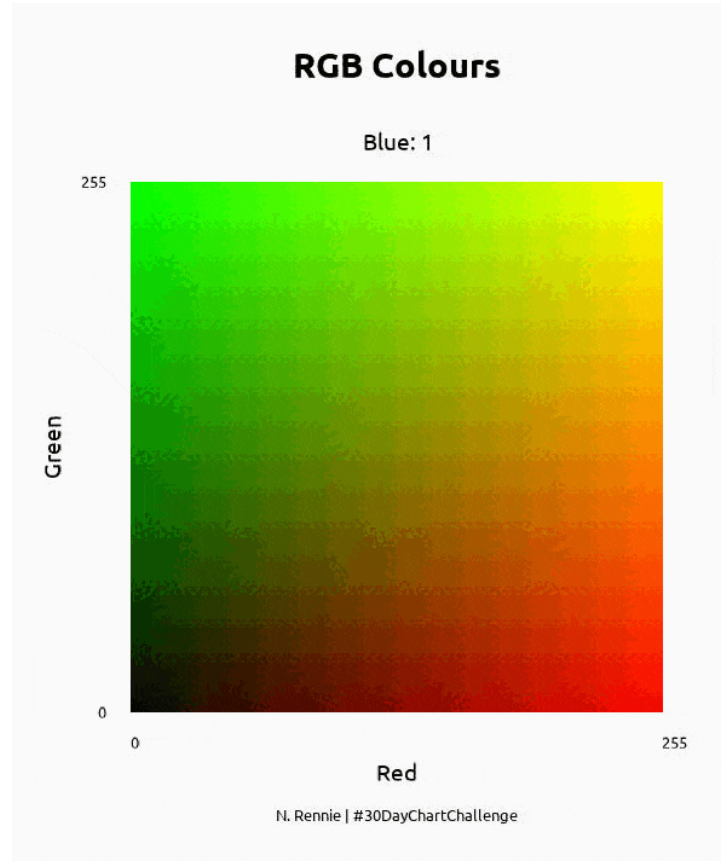
Nicola Rennie

Day 21 (Down and Upwards) in R



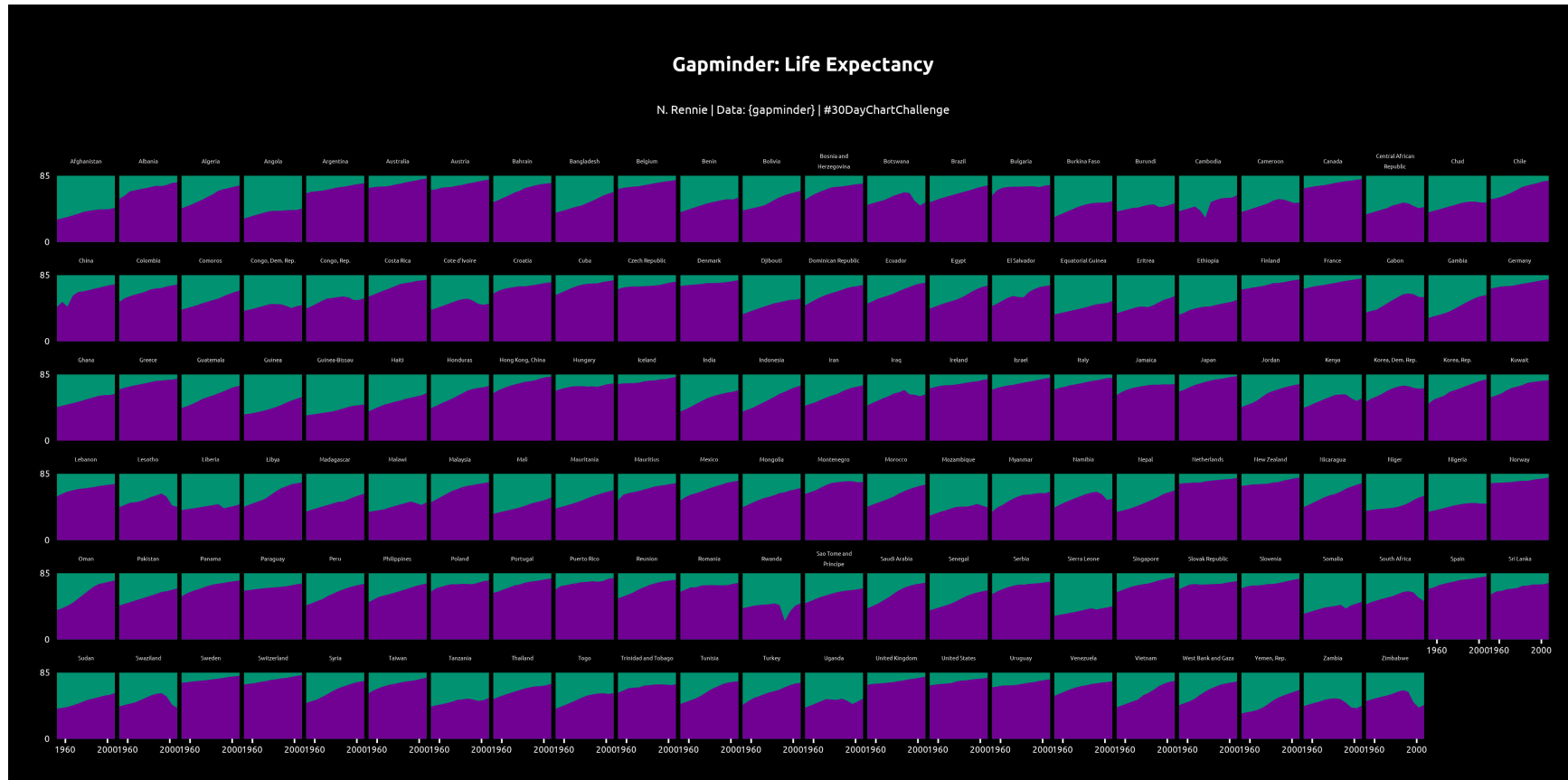
Nicola Rennie

Day 22 (Animation) in R



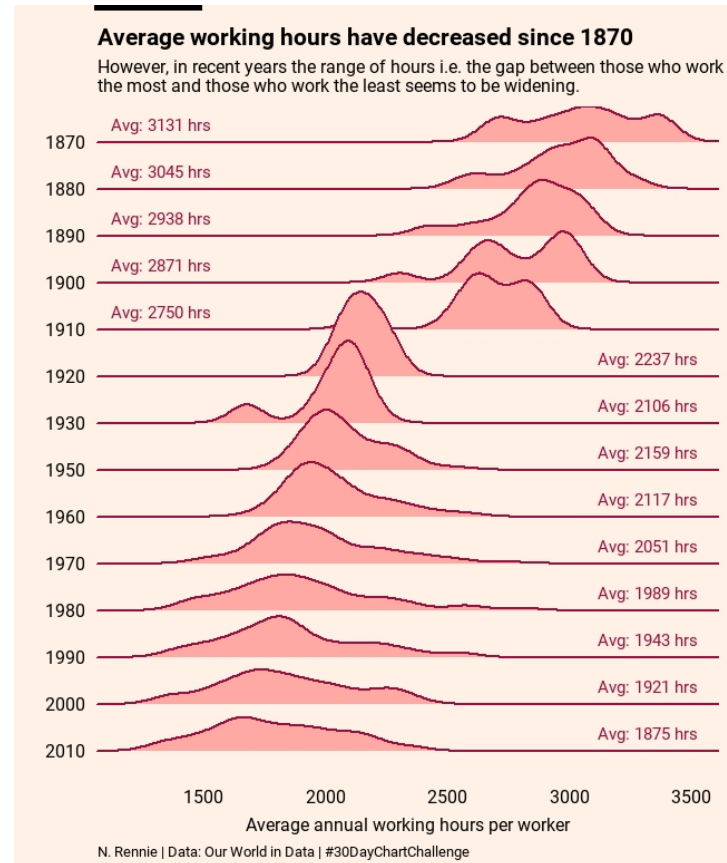
Nicola Rennie

Day 23 (Tiles) in R



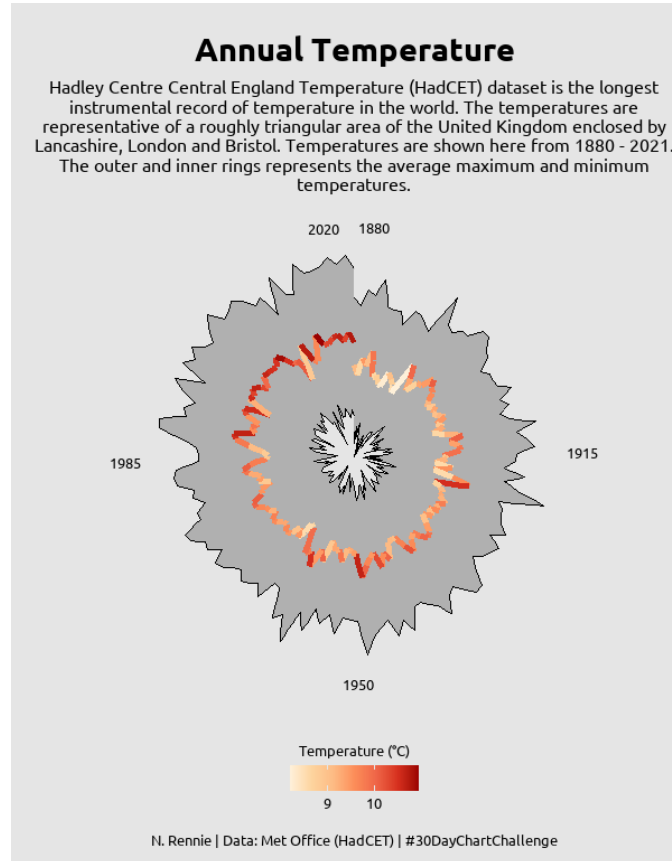
Nicola Rennie

Day 24 (Financial Times) in R



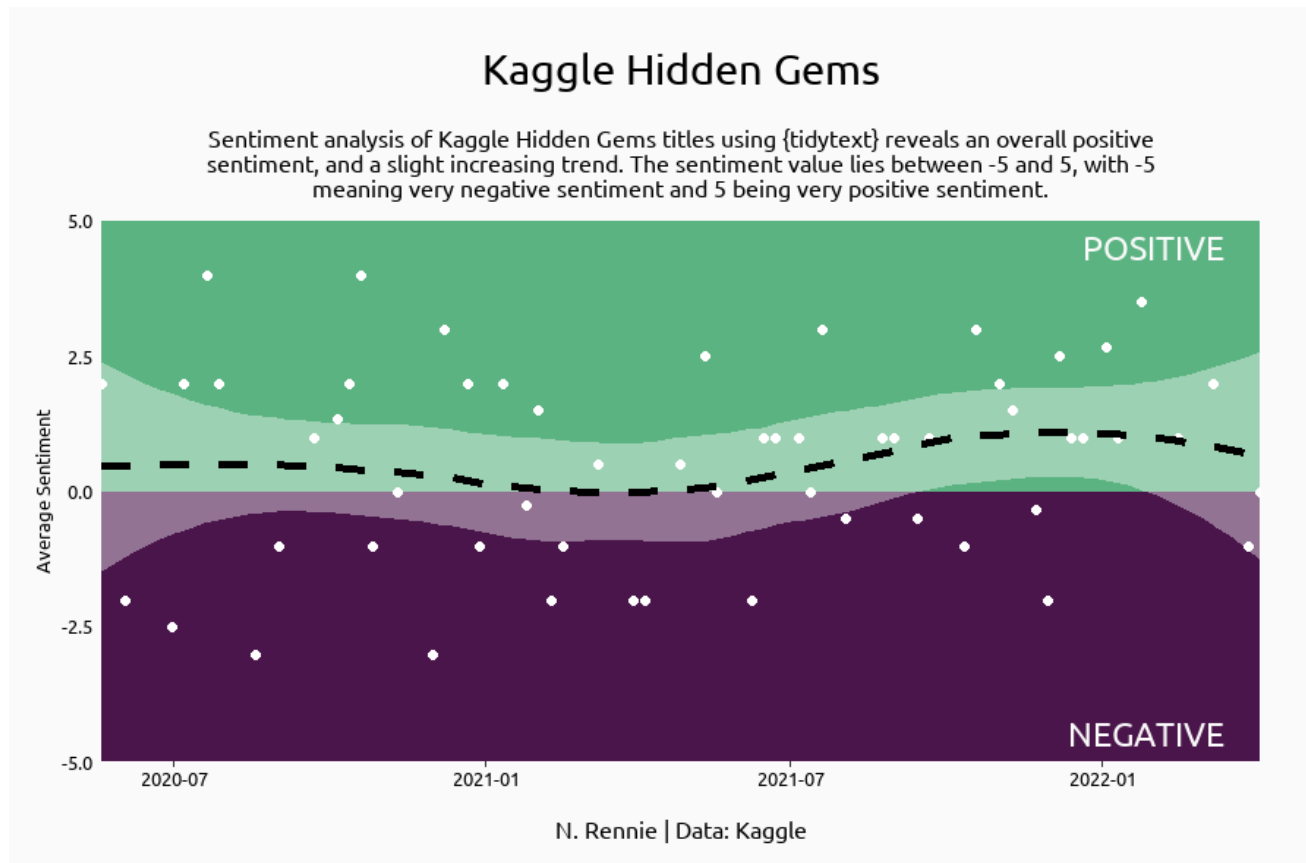
Nicola Rennie

Day 25 (Trend) in R



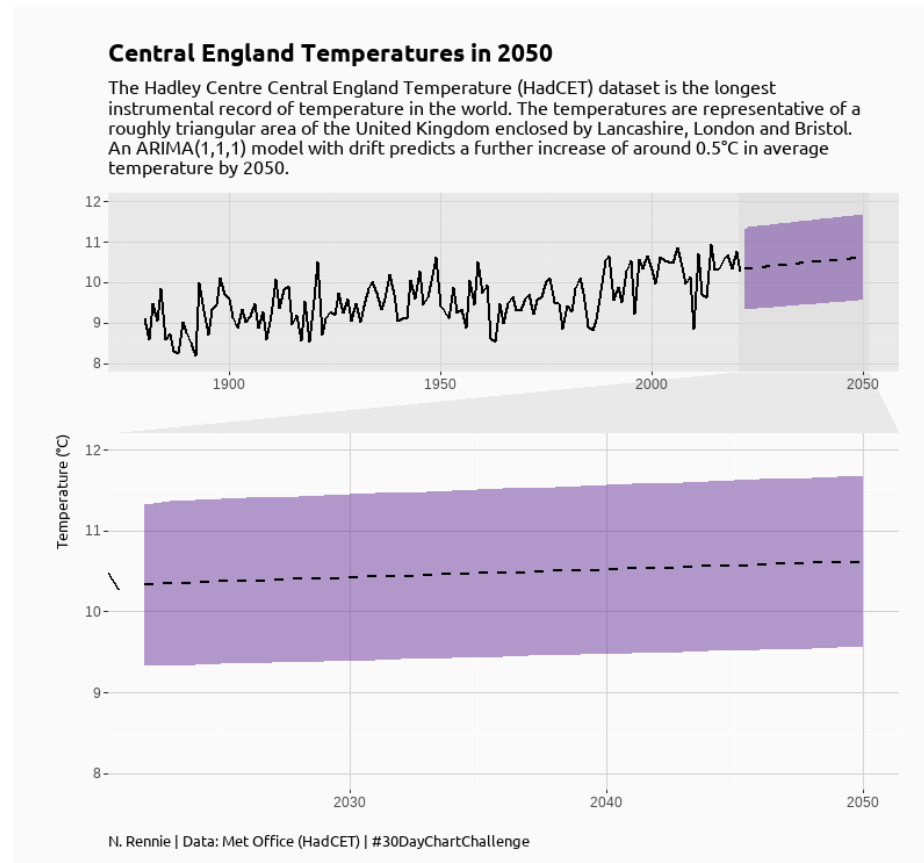
Nicola Rennie

Day 26 (Interactive) in R



Nicola Rennie

Day 27 (Future) in R

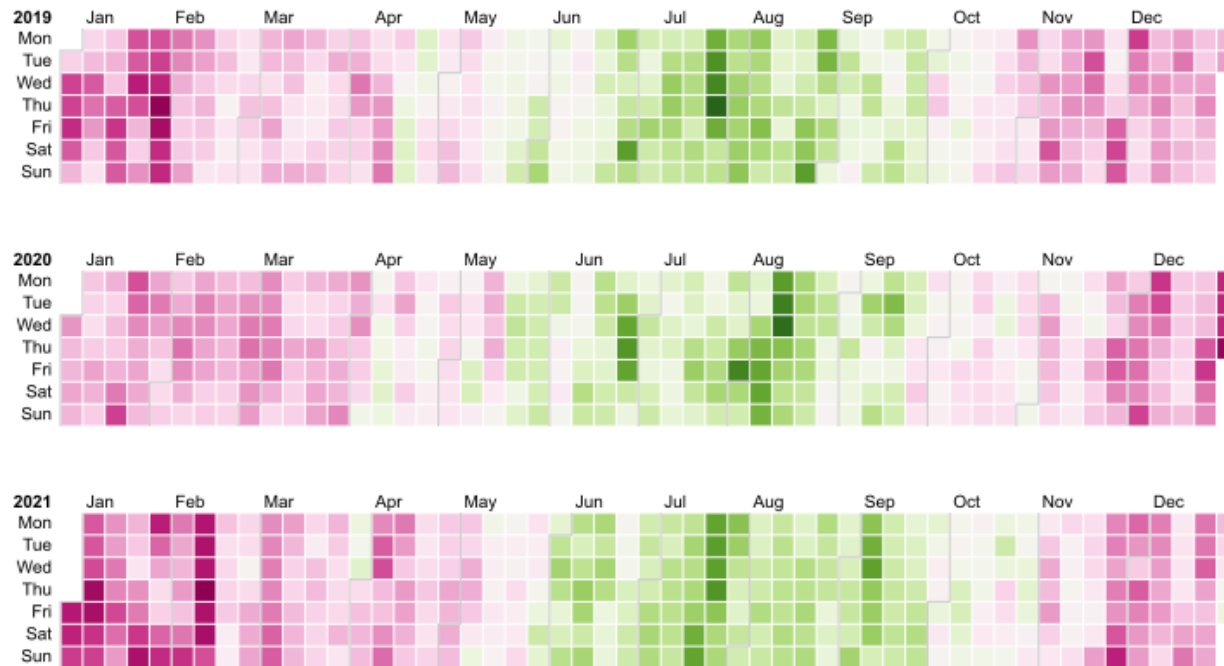


Nicola Rennie

Day 28 (Deviations) in RAWgraphs and Inkscape

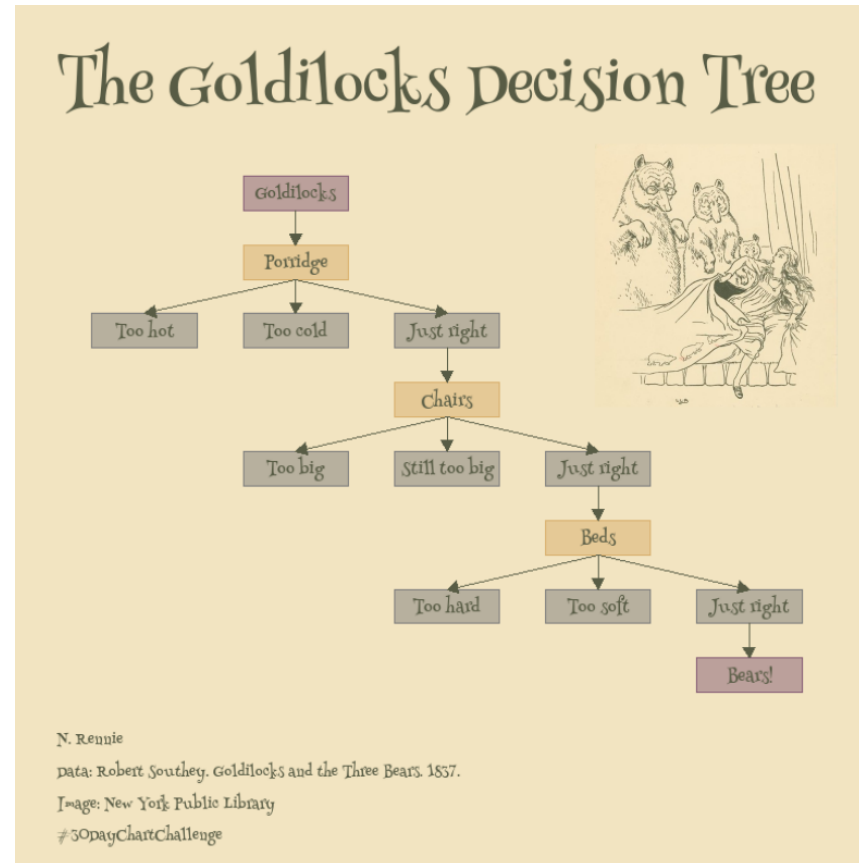
ANNUAL TEMPERATURE DEVIATIONS

N. Rennie | Data: Met Office (HadCET) | #30DayChartChallenge



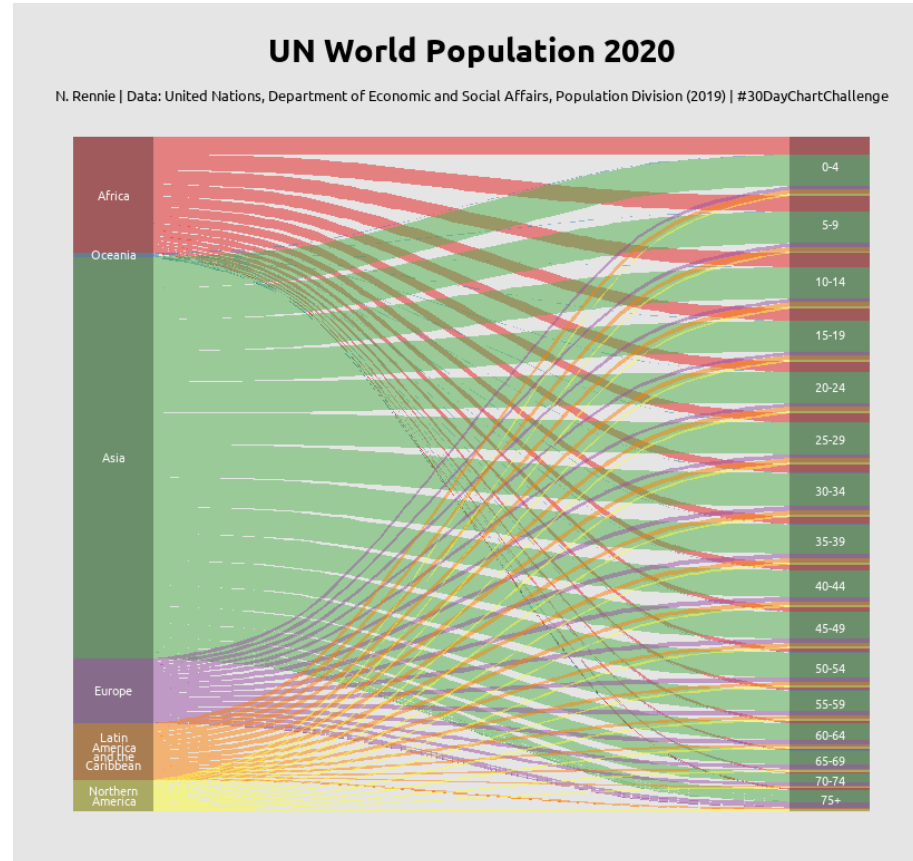
Nicola Rennie

Day 29 (Storytelling) in R and Inkscape



Nicola Rennie

Day 30 (UN Population) in R



Nicola Rennie

Lessons Learned

What did I learn?

- R packages
 - `facet_zoom()` from `{ggforce}`
 - Quarto
 - `{tidytext}`
- Non-R tools can be very helpful...
- Repeating styles should be bundled into an R package



What did I find difficult?

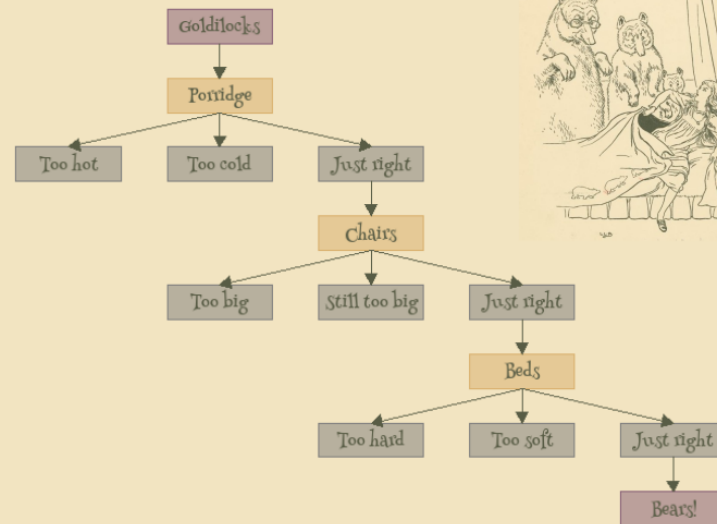
- Time
 - Didn't make charts each day, took breaks, reuse data
- Fitting my ideas and things I wanted to try to fit prompts
 - More planning



Part 2

The Goldilocks Decision Tree

The Goldilocks Decision Tree



N. Rennie

data: Robert Southey, Goldilocks and the Three Bears. 1837.

Image: New York Public Library

#30dayChartChallenge



Nicola Rennie

Flow charts in R

- `{grid}`
- `{DiagrammeR}`
- `{igraph}`
- `{ggnetwork}` / `{ggnet2}` / `{ggraph}`
- `{tikz}` (LaTeX)
- others...

Let's try to make a flowchart with `{ggplot2}`...



R Packages

```
library(tidyverse)  
library(igraph)  
library(showtext)  
library(rcartocolor)
```



Data

```
goldilocks <- tibble(from = c("Goldilocks",  
                              "Porridge", "Porridge", "Porridge",  
                              "Just right",  
                              "Chairs", "Chairs", "Chairs",  
                              "Just right2",  
                              "Beds", "Beds", "Beds",  
                              "Just right3"),  
  to = c("Porridge",  
         "Too cold", "Too hot", "Just right",  
         "Chairs",  
         "Still too big", "Too big", "Just right2",  
         "Beds",  
         "Too soft", "Too hard", "Just right3",  
         "Bears!"))
```



```
## # A tibble: 6 × 2
##   from      to
##   <chr>    <chr>
## 1 Goldilocks Porridge
## 2 Porridge   Too cold
## 3 Porridge   Too hot
## 4 Porridge   Just right
## 5 Just right Chairs
## 6 Chairs     Still too big
```



Nicola Rennie

Defining the layout

```
g = graph_from_data_frame(goldilocks, directed = TRUE)
coords = layout_as_tree(g)
colnames(coords) = c("x", "y")
```

```
##           x y
## [1,]    0 7
## [2,]    0 6
## [3,]   -1 5
## [4,]   -1 4
## [5,]   -2 3
## [6,]   -2 2
```



Adding attributes

```
output_df = as_tibble(coords) %>%  
  mutate(step = vertex_attr(g, "name"),  
         x = x*-1,  
         type = factor(c(1, 2, 3, 2, 3, 2, 3, 3, 3, 3, 3, 3, 3, 1)),  
         label = gsub("\\d+$", "", step))
```

```
## # A tibble: 6 × 5  
##       x     y step      type label  
##   <dbl> <dbl> <chr>    <fct> <chr>  
## 1     0     7 Goldilocks 1    Goldilocks  
## 2     0     6 Porridge  2    Porridge  
## 3     1     5 Just right 3    Just right  
## 4     1     4 Chairs    2    Chairs  
## 5     2     3 Just right2 3    Just right  
## 6     2     2 Beds      2    Beds
```



Making the boxes

```
plot_nodes = output_df %>%  
  mutate(xmin = x - 0.35,  
         xmax = x + 0.35,  
         ymin = y - 0.25,  
         ymax = y + 0.25)
```

```
## # A tibble: 6 × 9
```

##	x	y	step	type	label	xmin	xmax	ymin	ymax
##	<dbl>	<dbl>	<chr>	<fct>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	0	7	Goldilocks	1	Goldilocks	-0.35	0.35	6.75	7.25
## 2	0	6	Porridge	2	Porridge	-0.35	0.35	5.75	6.25
## 3	1	5	Just right	3	Just right	0.65	1.35	4.75	5.25
## 4	1	4	Chairs	2	Chairs	0.65	1.35	3.75	4.25
## 5	2	3	Just right2	3	Just right	1.65	2.35	2.75	3.25
## 6	2	2	Beds	2	Beds	1.65	2.35	1.75	2.25



Making the edges

```
plot_edges = goldilocks %>%  
  mutate(id = row_number()) %>%  
  pivot_longer(cols = c("from", "to"),  
               names_to = "s_e",  
               values_to = "step") %>%  
  left_join(plot_nodes, by = "step") %>%  
  select(-c(label, type, y, xmin, xmax)) %>%  
  mutate(y = ifelse(s_e == "from", ymin, ymax)) %>%  
  select(-c(ymin, ymax))
```

```
## # A tibble: 3 × 5  
##       id s_e  step          x      y  
##   <int> <chr> <chr>      <dbl> <dbl>  
## 1     1 1 from Goldilocks    0  6.75  
## 2     2 1 to   Porridge     0  6.25  
## 3     3 2 from Porridge     0  5.75
```



Choosing fonts

- Google fonts and the {showtext} package
- Browse fonts: fonts.google.com

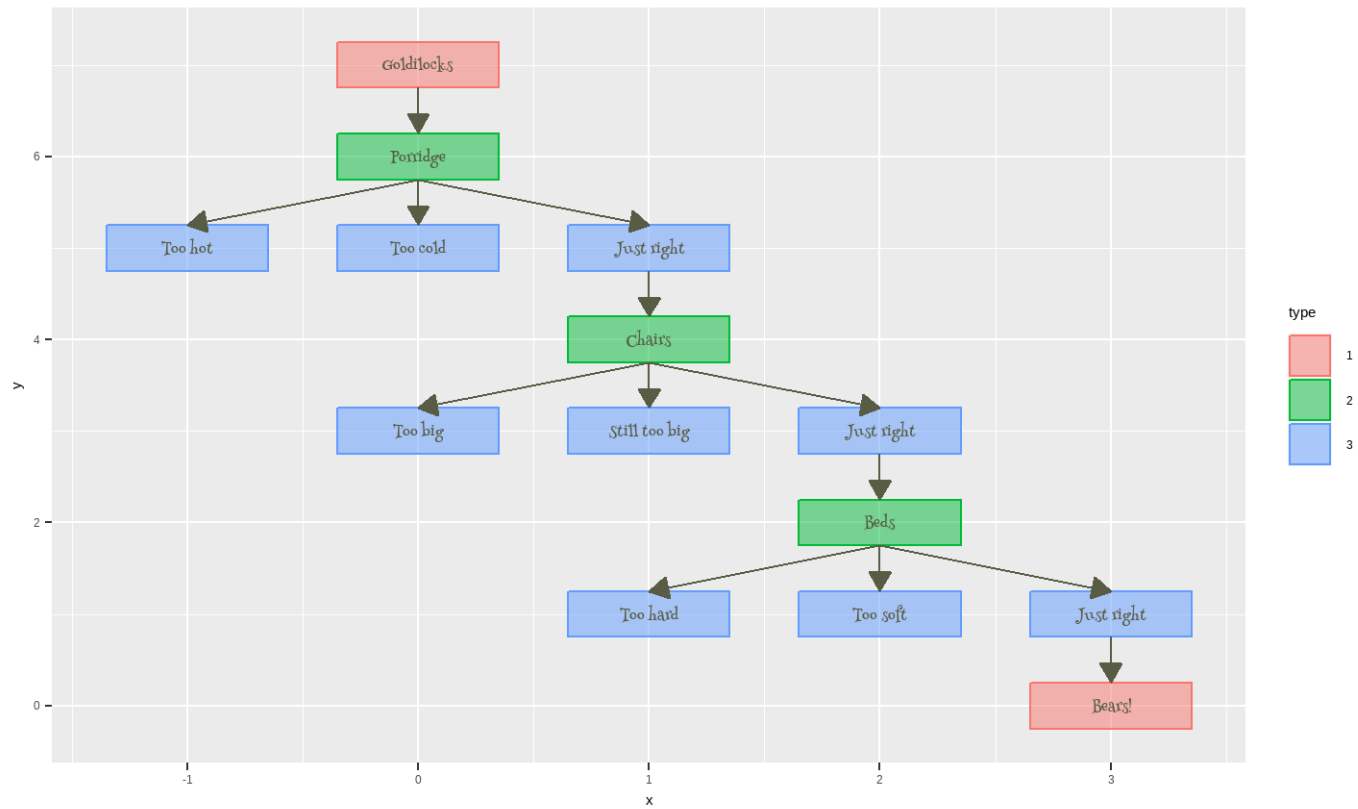
```
library(showtext)
font_add_google(name = "Henny Penny", family = "henny")
showtext_auto()
```



Plotting (finally!)

```
p = ggplot() +  
  # draw rectangles  
  geom_rect(data = plot_nodes,  
            mapping = aes(xmin = xmin, ymin = ymin, xmax = xmax, ymax = ymax,  
                          fill = type, colour = type),  
            alpha = 0.5,  
            linejoin = "round") +  
  # add text labels  
  geom_text(data = plot_nodes,  
            mapping = aes(x = x, y = y, label = label),  
            family = "henny",  
            color = "#585c45") +  
  # add arrows  
  geom_path(data = plot_edges,  
            mapping = aes(x = x, y = y, group = id),  
            colour = "#585c45",  
            arrow = arrow(length = unit(0.3, "cm"), type = "closed"))
```



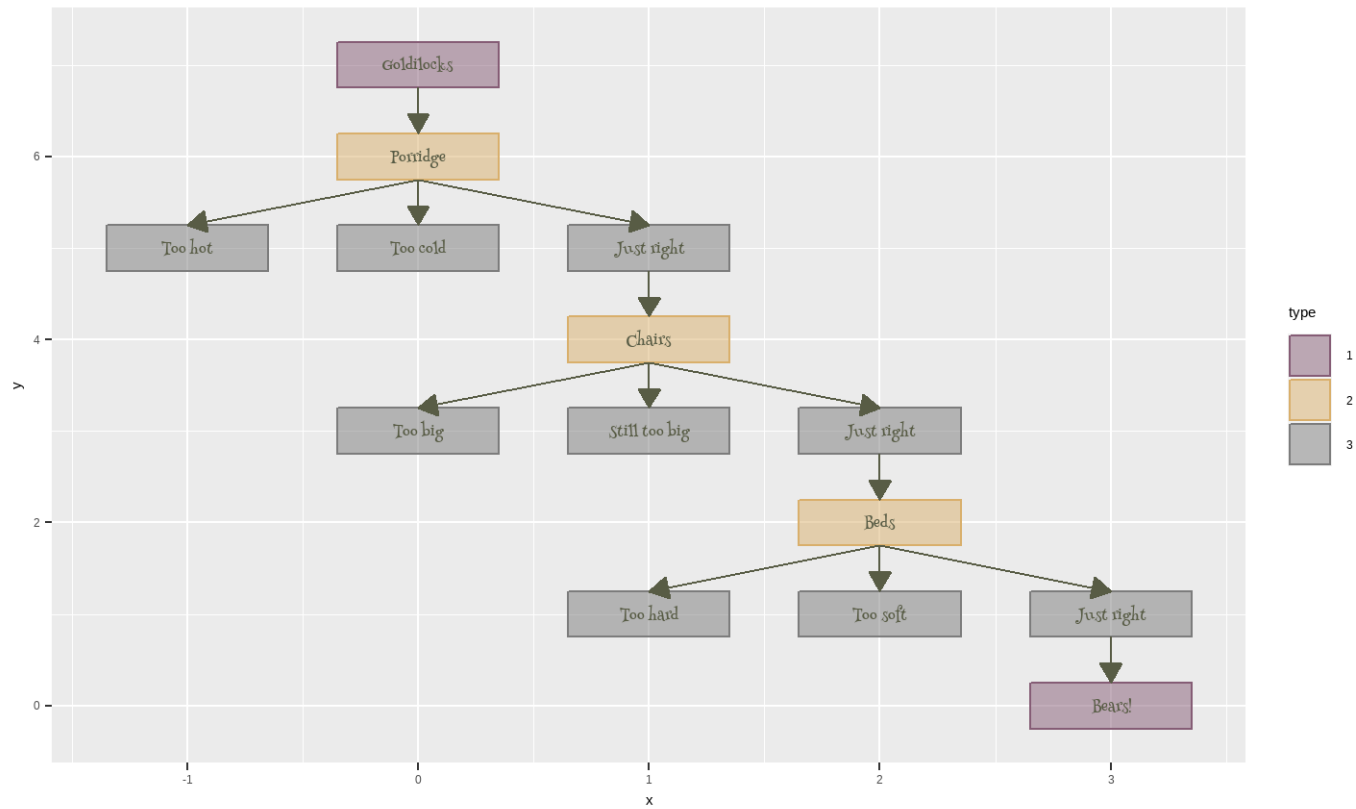


Colour schemes

- {rcartocolor}: jakubnowosad.com/rcartocolor

```
p = p +  
  scale_fill_carto_d(palette = "Antique") +  
  scale_colour_carto_d(palette = "Antique")
```



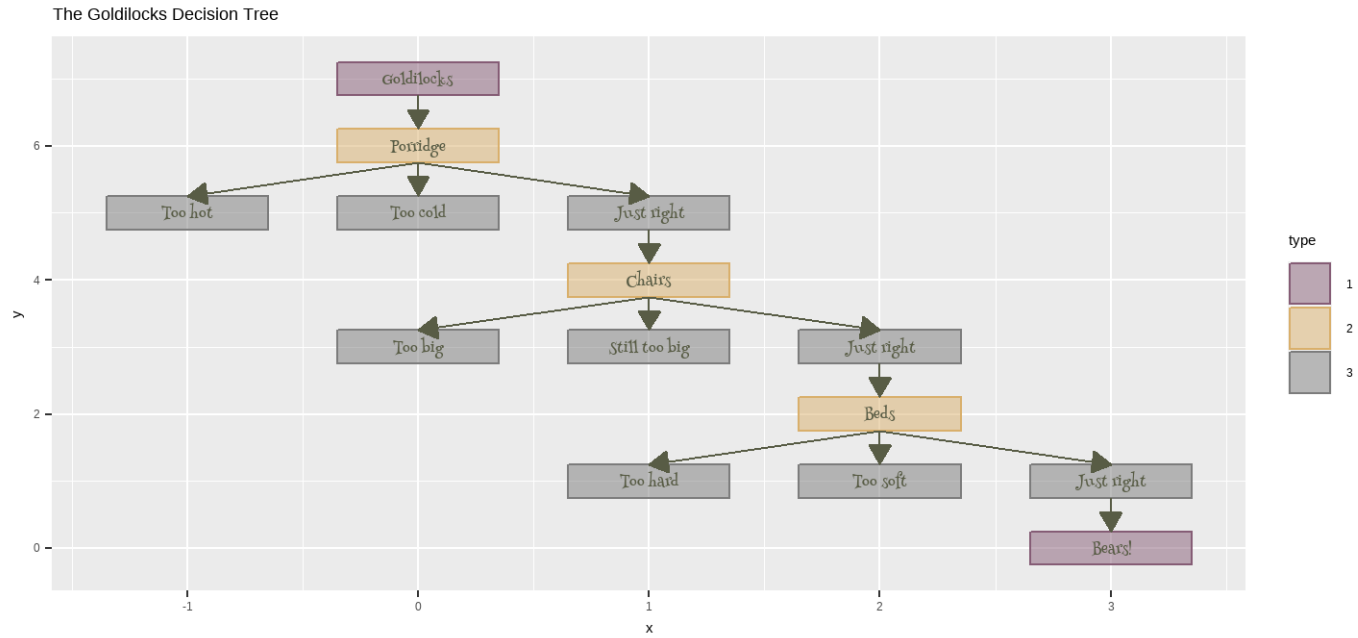


Some text labels

```
p = p +  
  labs(title = "The Goldilocks Decision Tree",  
        caption = "N. Rennie\n\nData: Robert Southey. Goldilocks and the Three Bears.  
1837.\n\nImage: New York Public Library\n\n#30DayChartChallenge")
```



Nicola Rennie



N. Rennie

Data: Robert Southey, Goldilocks and the Three Bears. 1837.

Image: New York Public Library

#30DayChartChallenge



Background colours

- Choose image
- Extract hex colour: imagecolorpicker.com

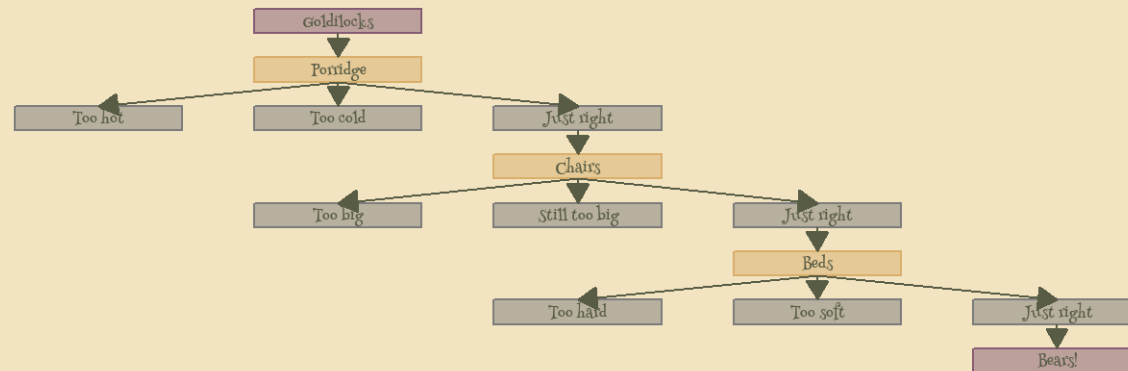


Themes

```
p = p +  
  theme_void() +  
  theme(plot.margin = unit(c(1, 1, 0.5, 1), "cm"),  
        legend.position = "none",  
        plot.background = element_rect(colour = "#f2e4c1", fill = "#f2e4c1"),  
        panel.background = element_rect(colour = "#f2e4c1", fill = "#f2e4c1"),  
        plot.title = element_text(family = "henny", hjust = 0, face = "bold",  
                                   size = 40, color = "#585c45",  
                                   margin = margin(t = 10, r = 0, b = 10, l = 0)),  
        plot.caption = element_text(family = "henny", hjust = 0,  
                                     size = 10, color = "#585c45",  
                                     margin = margin(t = 10)))
```



The Goldilocks Decision Tree



N. Rennie

Data: Robert Southey, Goldilocks and the Three Bears, 1837.

Image: New York Public Library

#OpagChartChallenge

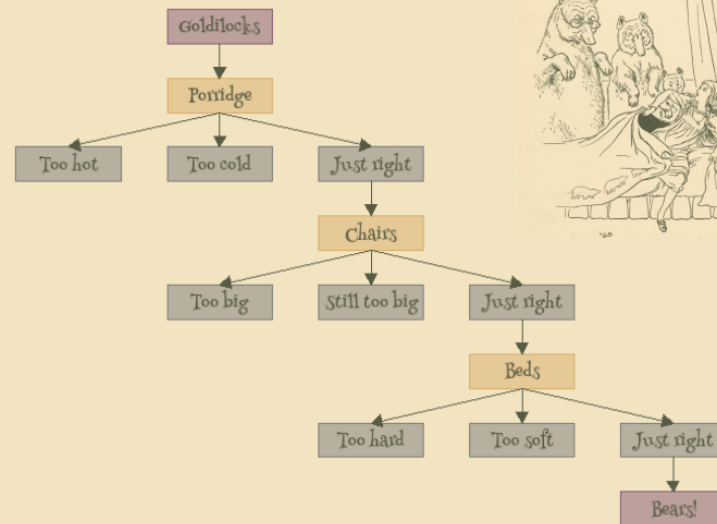


Adding images

- {magick} and {cowplot}
- Inkscape: inkscape.org



The Goldilocks Decision Tree



N. Rennie

data: Robert Southey, Goldilocks and the Three Bears. 1837.

Image: New York Public Library

#30dayChartChallenge



Nicola Rennie

- Twitter: [@nrennie35](https://twitter.com/nrennie35)
- GitHub: github.com/nrennie
- Website: nrennie.rbind.io
- Slides: nrennie.rbind.io/talks/2022-may-rladies-nairobi/

Questions?