

Exercise 9 | Grammar of Graphics II

Max Pellert (<https://mpellert.at>)

IS 616: Large Scale Data Analysis and Visualization

Du Bois

William Edward Burghardt Du Bois ([/dju:'bɔɪs/ dew-BOYSS](#);^{[1][2]} February 23, 1868 – August 27, 1963) was an American [sociologist](#), [socialist](#), historian, and [Pan-Africanist civil rights activist](#). Born in [Great Barrington, Massachusetts](#), Du Bois grew up in a relatively tolerant and [integrated](#) community. After completing graduate work at the [Friedrich Wilhelm University](#) (in [Berlin](#), Germany) and [Harvard University](#), where he was the first African American to earn a doctorate, he became a professor of history, sociology, and economics at [Atlanta University](#). Du Bois was one of the founders of the [National Association for the Advancement of Colored People](#) (NAACP) in 1909.

His 1940 autobiography [Dusk of Dawn](#) is regarded in part as one of the first scientific treatises in the field of American sociology, and he published two other life stories, all three containing essays on sociology, politics and history. In his role as editor of the NAACP's journal [The Crisis](#), he published many influential pieces. Du Bois believed that [capitalism](#) was a primary cause of racism, and he was generally sympathetic to [socialist](#) causes throughout his life. He was an ardent peace activist and advocated [nuclear disarmament](#). The United States [Civil Rights Act](#), embodying many of the reforms for which Du Bois had campaigned his entire life, was enacted a year after his death.



1900 Paris Exposition

Du Bois was the primary organizer of *The Exhibit of American Negroes* at the *Exposition Universelle* held in Paris between April and November 1900, for which he put together a series of 363 photographs aiming to commemorate the lives of African Americans at the turn of the century and challenge the racist caricatures and stereotypes of the day.^{[58][59]} Also included were charts, graphs, and maps.^{[60][61]} He was awarded a gold medal for his role as compiler of the materials, which are now housed at the [Library of Congress](#).^[59]

https://en.wikipedia.org/wiki/W._E._B._Du_Bois

The Exhibit of American Negroes was a sociological display within the Palace of Social Economy at the [1900 World's Fair](#) in [Paris](#). The exhibit was a joint effort between [Daniel Murray](#), the Assistant Librarian of Congress, [Thomas J. Calloway](#), a lawyer and the primary organizer of the exhibit, and [W. E. B. Du Bois](#). The goal of the exhibition was to demonstrate progress and commemorate the lives of [African Americans](#) at the turn of the century.^[1]

The exhibit included a statuette of [Frederick Douglass](#), four bound volumes of nearly 400 official [patents](#) by African Americans, photographs from several educational institutions ([Fisk University](#), [Howard University](#), [Roger Williams University](#), [Tuskegee Institute](#), [Claflin University](#), [Berea College](#), [North Carolina A&T](#)), an African-American [bibliography](#) by the [Library of Congress](#) containing 1,400 titles, and two social studies directed by W. E. B. Du Bois: "The Georgia Negro" (comprising 32 handmade graphs and charts), and a set of about 30 statistical graphics on the African-American population made by Du Bois's students at Atlanta University.^[2] Most memorably, the exhibit displayed some five hundred photographs of African-American men and women, homes, churches, businesses and landscapes including photographs from [Thomas E. Askew](#).

https://en.wikipedia.org/wiki/The_Exhibit_of_American_Negroes



EXHIBIT OF AMERICAN NEGROES AT THE PARIS EXPOSITION.

Renewed interest in Du Bois' forgotten visualizations

HISTORY

Document Deep Dive

W.E.B. Du Bois' Visionary Infographics Come Together for the First Time in Full Color

His pioneering team of black sociologists created data visualizations that explained institutionalized racism to the world

Jackie Mansky

November 15, 2018

<https://www.smithsonianmag.com/history/first-time-together-and-color-book-displays-web-du-bois-visionary-infographics-180970826/>

Relevance today

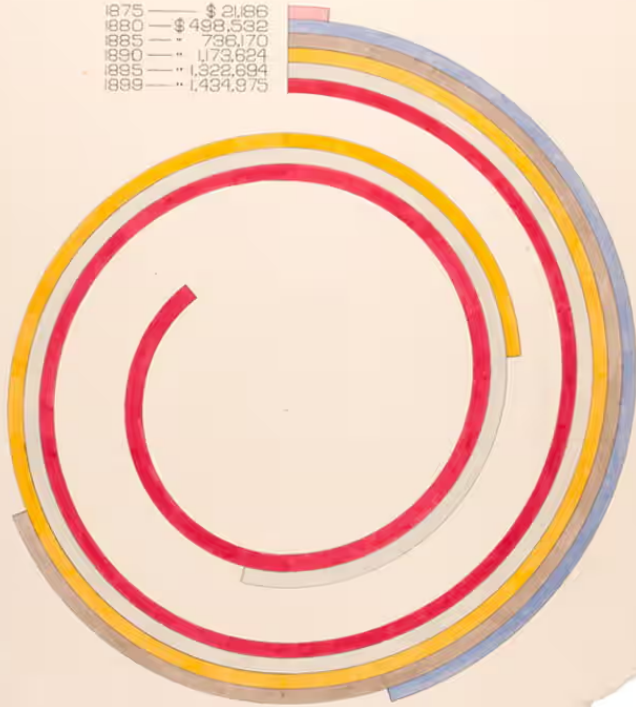
WEB Du Bois: retracing his attempt to challenge racism with data

The civil rights pioneer and scholar is most famous for his book *The Souls of Black Folk*, but his use of data to show inequality is still profound today

<https://www.theguardian.com/world/2017/feb/14/web-du-bois-racism-data-paris-african-americans-jobs>

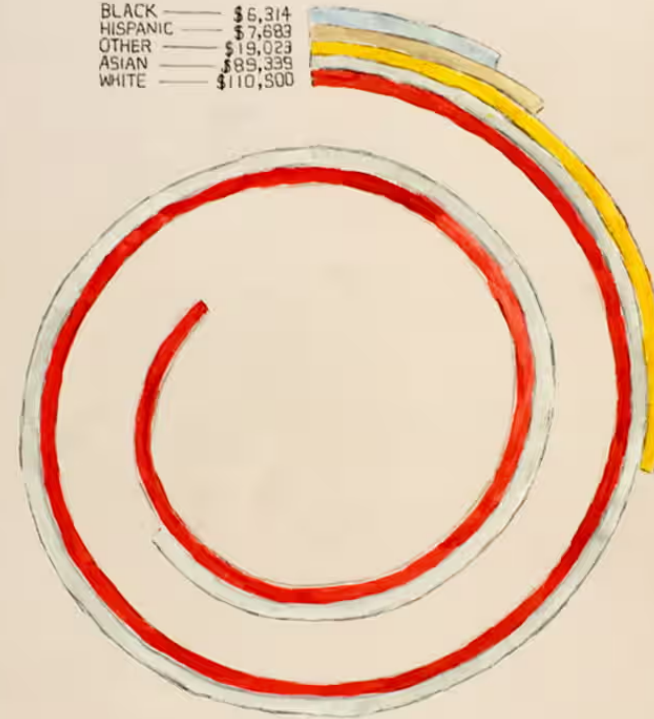
ASSESSED VALUE OF HOUSEHOLD AND KITCHEN FURNITURE OWNED BY GEORGIA NEGROES.

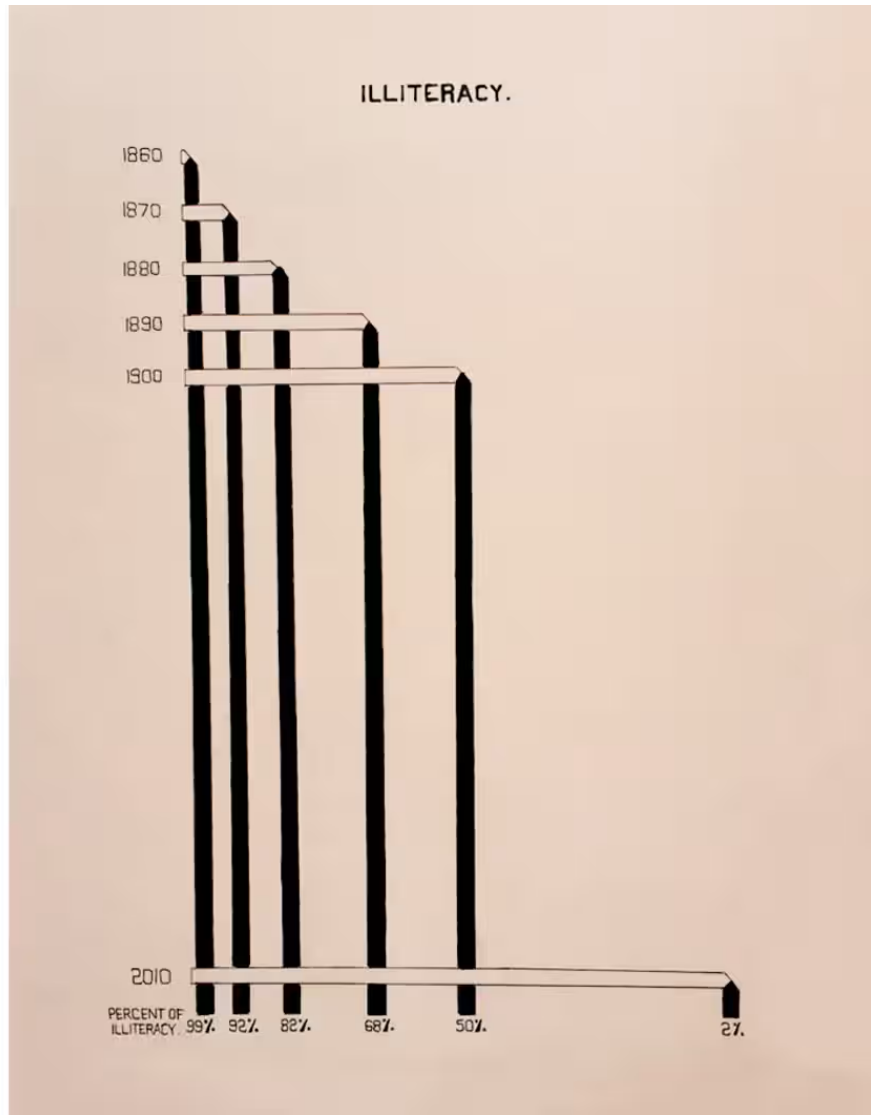
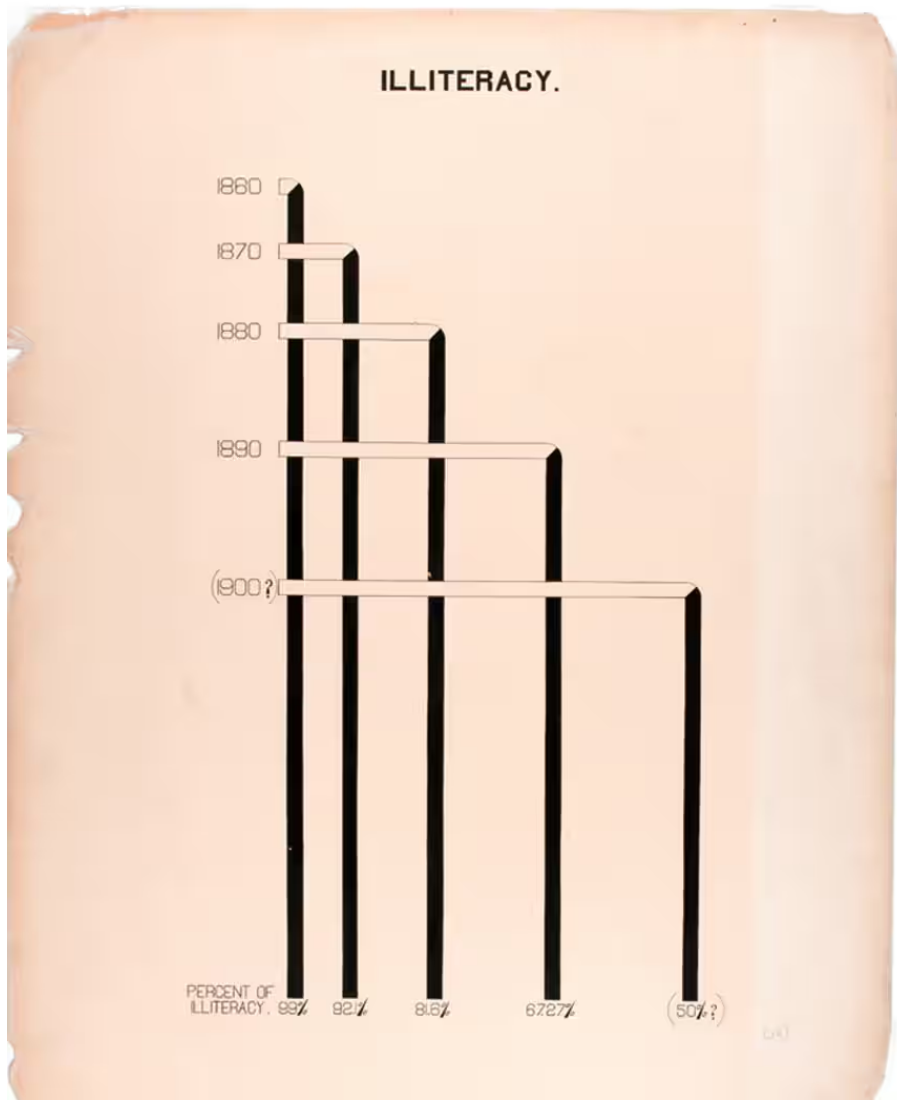
1875	—	\$ 21,866
1880	—	\$ 498,532
1885	—	796,170
1890	—	1,173,624
1895	—	1,322,694
1899	—	1,434,975



NET WORTH OF HOUSEHOLDS

BLACK	—	\$ 6,314
HISPANIC	—	\$ 7,683
OTHER	—	\$ 19,023
ASIAN	—	\$ 89,339
WHITE	—	\$ 110,500





Six-part series on Du Bois' visualizations

I. [The Exhibit of American Negroes](#) An introduction to the 1900 Paris Exposition which discusses a few notable charts that focus on history and population growth.

II. [Data Journalism and the Scientific Study of "The Negro Problem"](#) Places this body of work within Du Bois' larger sociological focus and continues the exploration of many of the charts from the exposition with a focus on education, literacy, and occupation.

III. [Exploring the Craft and Design of W.E.B. Du Bois' Data Visualizations](#) A detailed examination on how Du Bois drafted his charts, a consideration of this work as a precursor to modernism, and the discussion of a series of charts on land ownership and value.

W. E. B. Du Bois' staggering Data Visualizations are as powerful today as they were in 1900 (Part 1)

Six-part series on Du Bois' visualizations

IV. *Style and Rich Detail; On Viewing an Original Du Bois Chart* Discoveries on viewing an original chart and further exploration of Du Bois' more innovative charts dealing with occupation, business, and mortality.

V. *Du Bois as Social Scientist and the Legacy of "The Exhibit of American Negroes"* Will Discuss Du Bois' body of work from this period and his frustrations with social science despite widespread attention.

VI. *The Exhibition as a Whole: an Exciting Discovery* To close out the series I'll present a very exciting discovery I've made and will present each chart in sequence."

For all the original files: <https://www.loc.gov/pictures/search/?sp=1&co=anedub&st=grid>

Data Journalism in the study of W.E.B. Du Bois' "The Negro Problem" (Part 2)

Du Bois Style

“Circles

Several charts use circular elements; notable are the spirals in plates 11 and 25 (often highlighted when showing the Du Bois visuals).

The spirals are used to indicate large measures; instead of stretching out the lines as in a conventional bar chart, the measures are rolled up in a spiral.”

<https://github.com/ajstarks/dubois-data-portraits/blob/master/dubois-style.pdf>

**Let's think a bit how to do this
in R**

```

1 library(tidyverse)
2 library(scales)
3 library(extrafont)
4
5 extrafont::loadfonts(
6   device = 'win')
7
8 # As a minimal working example,
9 # here's some toy data
10
11 df <- tibble(
12   x = rep(c(0, 10), 5),
13   y = rep(10:5, each = 2)[2:11],
14   g = rep(1:5, each = 2)
15 )
16
17 df

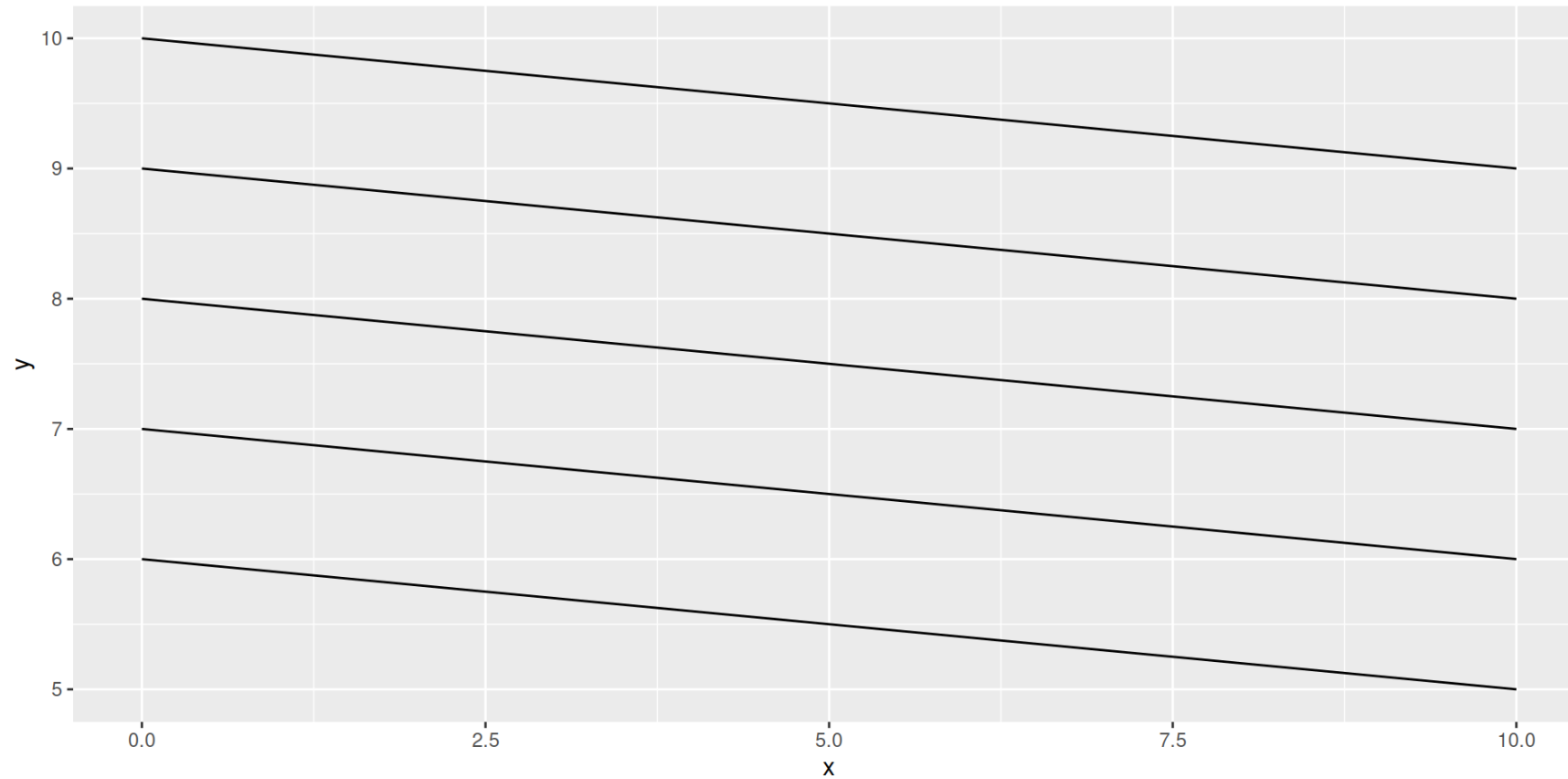
```

```

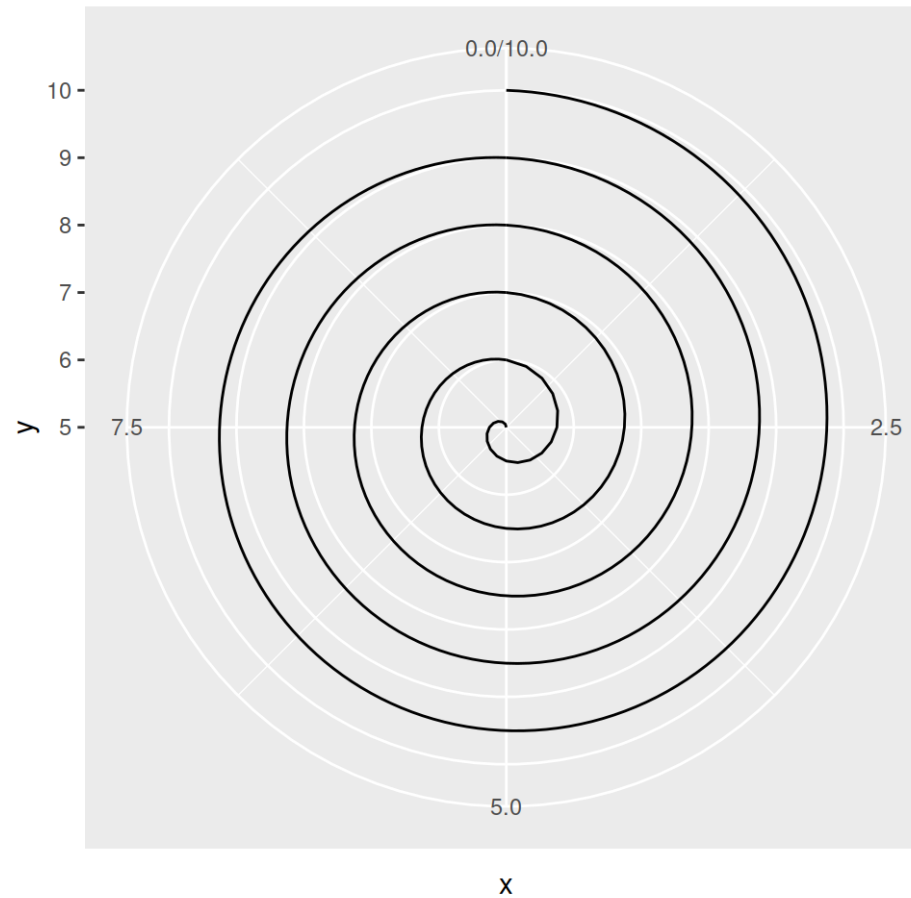
# A tibble: 10 × 3
      x     y     g
  <dbl> <int> <int>
1     0    10     1
2    10     9     1
3     0     9     2
4    10     8     2
5     0     8     3
6    10     7     3
7     0     7     4
8    10     6     4
9     0     6     5
10    10     5     5

```

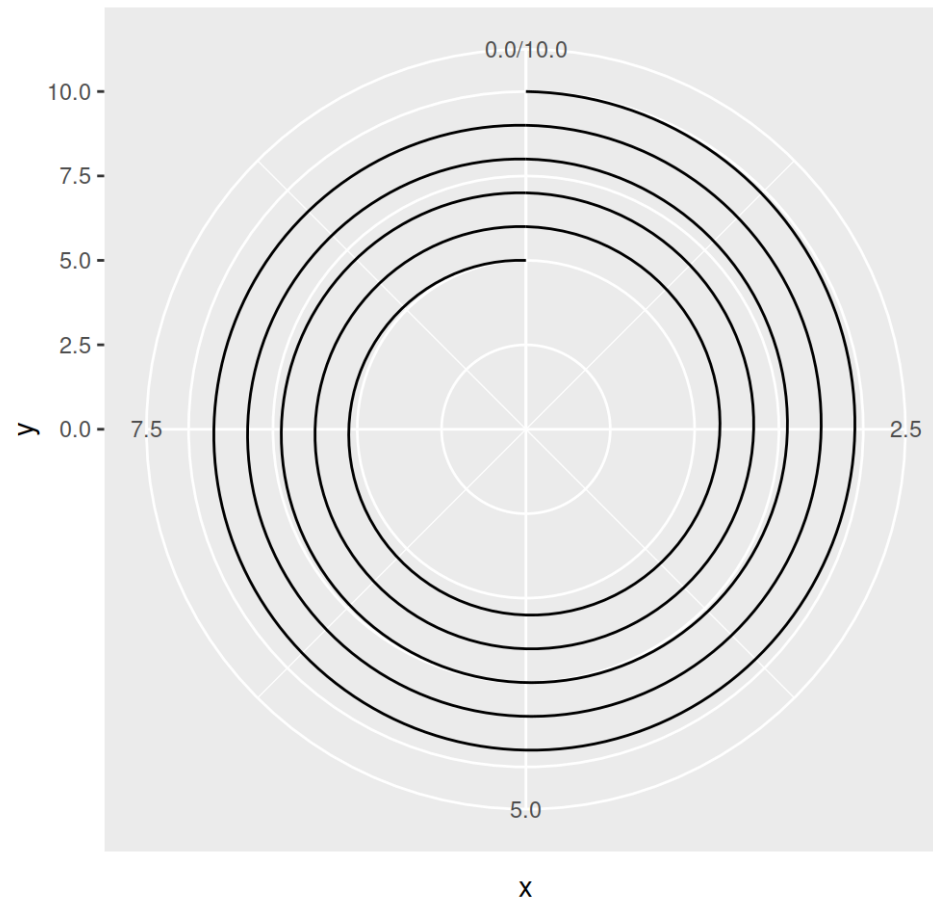
```
1 library(ggplot2)
2
3 ggplot(df) +
4   aes(x = x,
5       y = y,
6       group = g) +
7   geom_line()
```



```
1 t1 <- ggplot(df) +  
2   aes(x = x, y = y,  
3       group = g) +  
4   geom_line() +  
5   coord_polar()
```

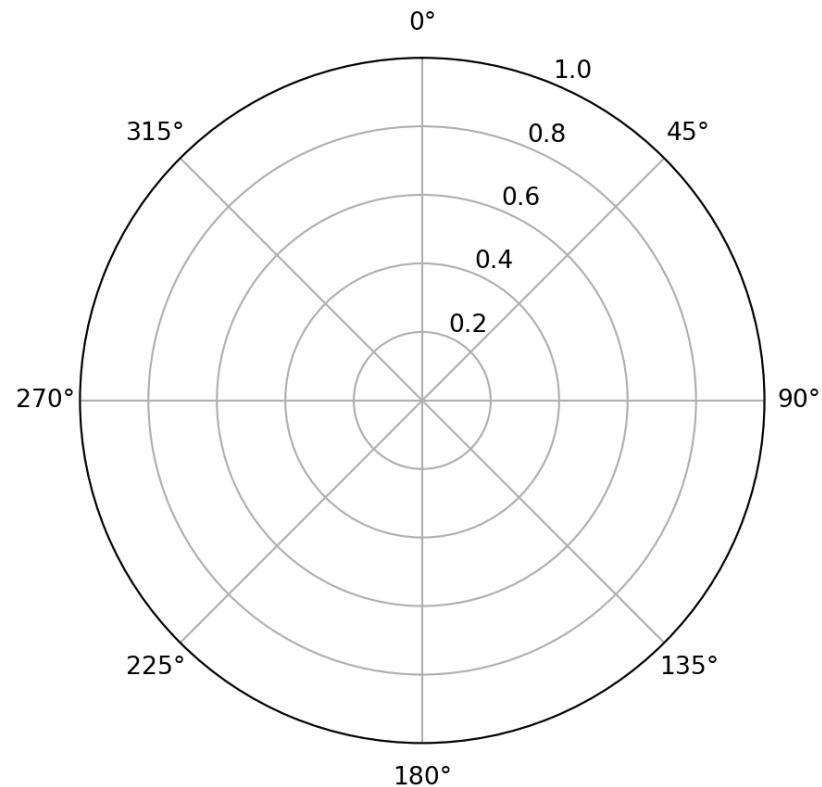



```
1 # By adjusting the limits
2 # of the y-axis,
3 # you can control the density
4 # of the spiral
5
6 t1 + ylim(c(0, 10))
```

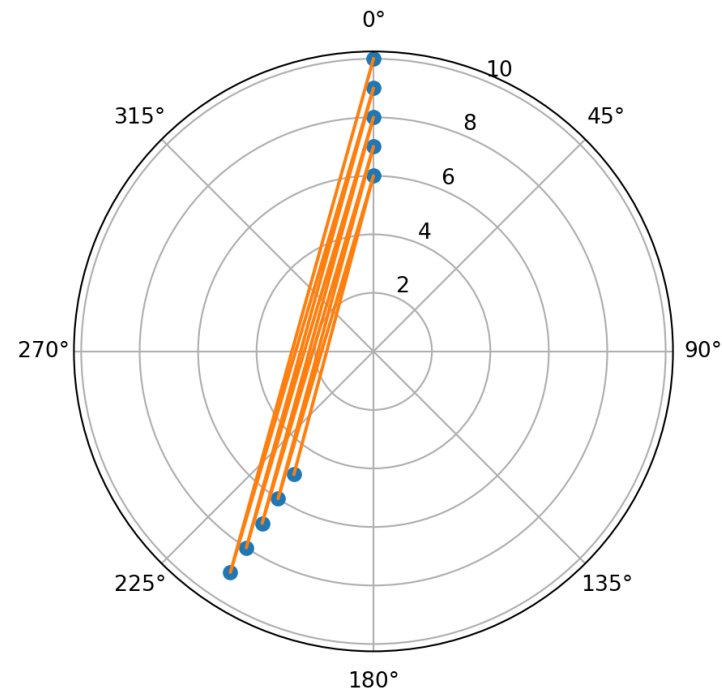


Let's try that in Python

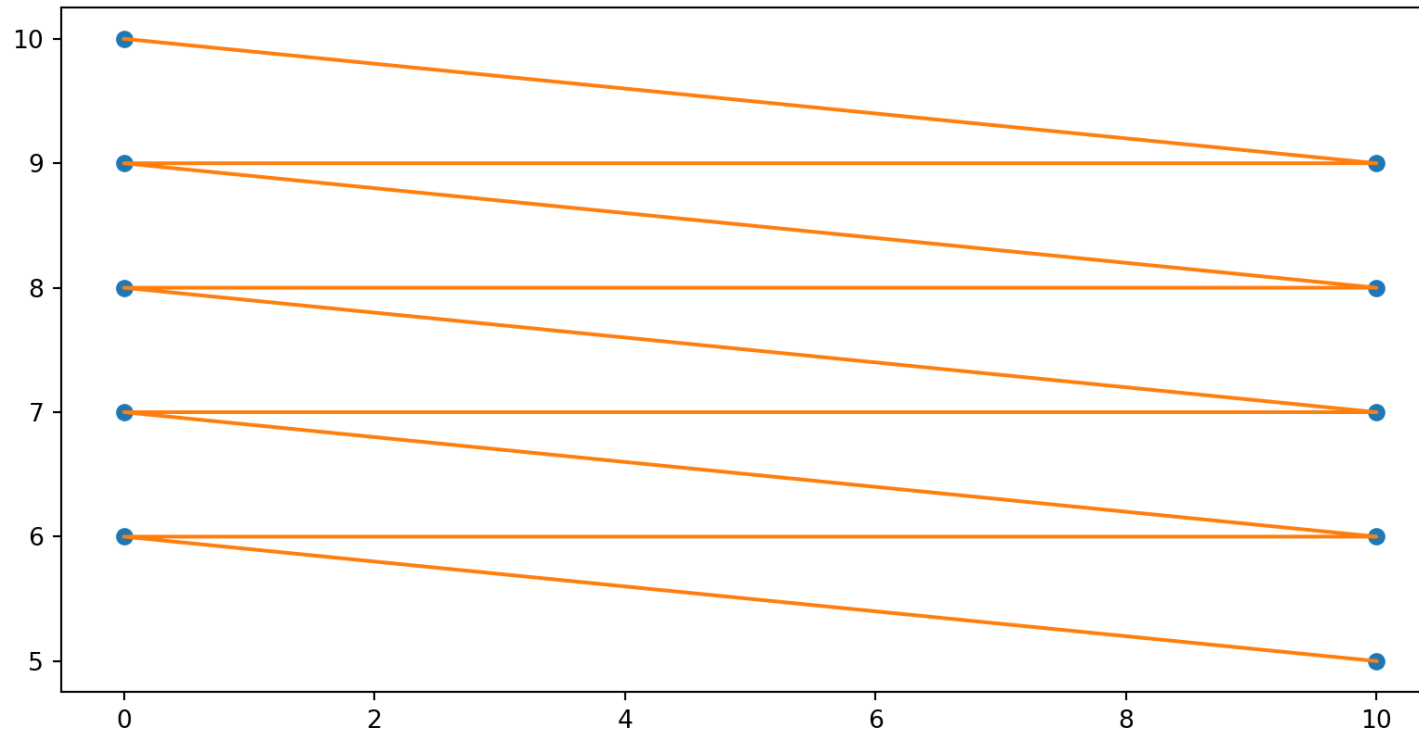
```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 import pandas as pd
4
5 fig, ax = plt.subplots(subplot_kw={'projection': 'polar'})
6 ax.set_theta_zero_location("N")
7 ax.set_theta_direction(-1)
8 plt.show()
```



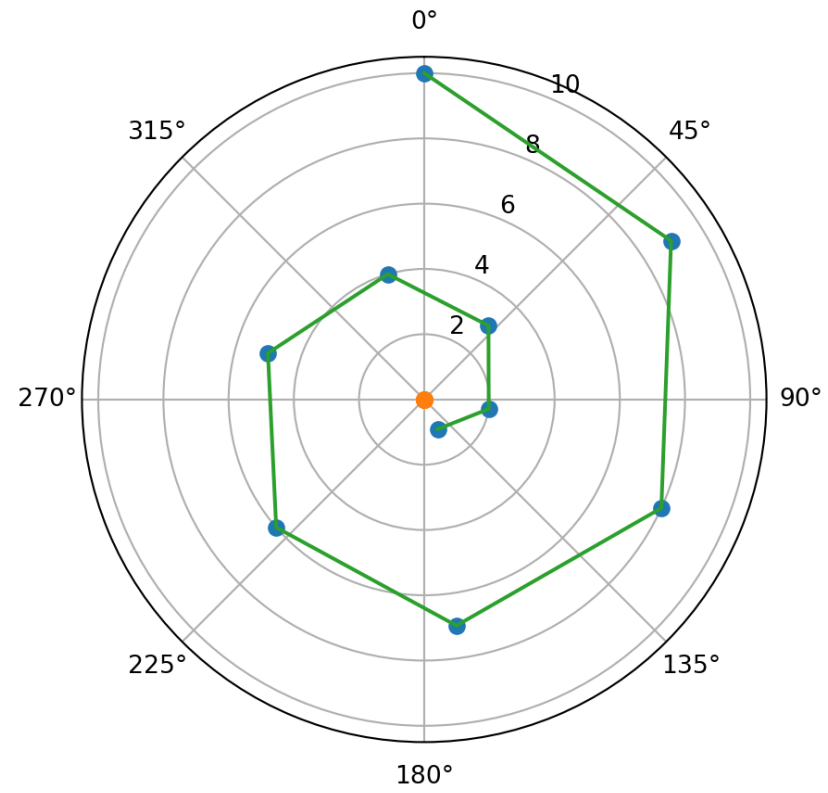
```
1 x = [0, 10, 0, 10, 0, 10, 0, 10, 0, 10]
2 y = [10, 9, 9, 8, 8, 7, 7, 6, 6, 5]
3
4 fig, ax = plt.subplots(subplot_kw={'projection': 'polar'})
5 ax.plot(x,y, "o")
6 ax.plot(x,y)
7 ax.set_theta_zero_location("N")
8 ax.set_theta_direction(-1)
9 plt.show()
```




```
1 fig, ax = plt.subplots()
2 ax.plot(x,y, "o")
3 ax.plot(x,y)
4 plt.show()
```



```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 import pandas as pd
4
5 def cart2polNP(xy):
6     x, y = xy
7     rho = np.sqrt(x**2 + y**2)
8     phi = np.arctan2(y, x)
9     return(rho, phi)
10
11 x = [10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
12 y = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
13
14 pol=[cart2polNP(x) for x in zip(x,y)]
15
16 fig, ax = plt.subplots(subplot_kw={'projection': 'polar'})
17 ax.set_theta_zero_location("N")
18 ax.set_theta_direction(-1)
19 ax.plot(pol, "o")
20 ax.plot(pol)
21 plt.show()
```



I have a solution that is a patchwork of other solutions. It needs to be cleaned and optimized, but it does the job !

Comments and improvements are always welcome, see below

```
# https://stackoverflow.com/questions/33962717/interpolating-a-closed-curve-using-  
  
from scipy import interpolate  
  
x=df.measure[:-1] * np.cos(df.angle[:-1]/180*np.pi)  
y=df.measure[:-1] * np.sin(df.angle[:-1]/180*np.pi)  
x = np.r_[x, x[0]]  
y = np.r_[y, y[0]]  
  
# fit splines to x=f(u) and y=g(u), treating both as periodic. also note that s=0  
# is needed in order to force the spline fit to pass through all the input points.  
tck, u = interpolate.splprep([x, y], s=0, per=True)  
  
# evaluate the spline fits for 1000 evenly spaced distance values  
xi, yi = interpolate.splev(np.linspace(0, 1, 1000), tck)  
  
def cart2pol(x, y):  
    rho = np.sqrt(x**2 + y**2)  
    phi = np.arctan2(y, x)  
    return(rho, phi)
```

<https://stackoverflow.com/questions/59493724/custom-spider-chart-display-curves-instead-of-lines-between-point-on-a-polar>

Let's return to R...

All credits to https://twitter.com/ijeamaka_a/status/1361715338027560962 &
https://github.com/Ijeamakaanyene/tidytuesday/blob/master/scripts/2021_06_dubois_data.Rmd

```
1 library(ggplot2)
2 library(dplyr)
3 library(glue)
4
5 extrafont::loadfonts()
6
7 furniture = readr::read_csv('https://raw.githubusercontent.com/rfordatascie
8   janitor::clean_names()
9
10 # Color palette pulled from original image
11 bcgrnd = "#e4d2c1"
12 yr_1875 = "#eaafa6"
13 yr_1880 = "#9da0b0"
14 yr_1885 = "#c4a58f"
15 yr_1890 = "#ecb025"
16 yr_1895 = "#d8c7b3"
17 yr_1899 = "#dc354a"
18
19 dubois_palette = c(`1875` = yr_1875,
20                   `1880` = yr_1880,
21                   `1885` = yr_1885,
22                   `1890` = yr_1890,
```

```

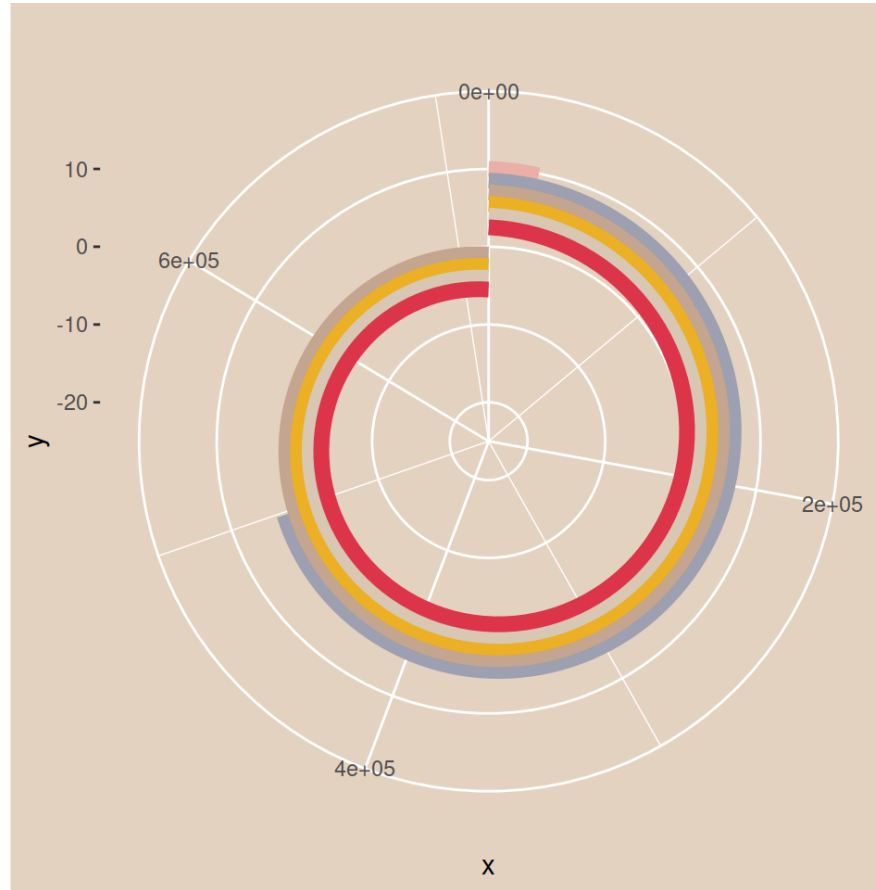
1 max_x = max(furniture$household_value_dollars) / 2
2
3 # Slope calculated using year 1885
4 slope = (-1 - 7) / (717487.5 - 0)
5
6 furniture_spiral = furniture %>%
7   mutate(year = as.factor(year),
8          y = seq(10, by = -1.5, length.out = 6),
9          x = 0) %>%
10  rowwise() %>%
11  mutate(xend = min(household_value_dollars, max_x),
12         yend = slope*xend + y) %>%
13  mutate(y_2 = yend,
14         x_2 = 0,
15         xend_2 = if_else(household_value_dollars < max_x,
16                          NA_real_, household_value_dollars - max_x),
17         yend_2 = slope*xend_2 + y_2)

```

```

1  ggplot(data = furniture_spiral) +
2    # first part of spiral
3    geom_segment(aes(x = x, xend = xend,
4                    y = y, yend = yend,
5                    color = year),
6                size = 3) +
7    coord_polar(clip = "off") +
8    ylim(-25, 15) +
9    xlim(0, 717487.5) +
10   scale_color_manual(values = dubois_palette) +
11   theme(plot.background = element_rect(fill = bcgrnd,
12                                       color = NA),
13         plot.margin = margin(t = 20, r = 5, b = 5, l = 5),
14         panel.background = element_rect(fill = bcgrnd,
15                                         color = NA),
16         plot.title = element_text(hjust = 0.5, family = "Cutive",
17                                   face = "bold", size = 15),
18         plot.caption = element_text(size = 4, family = "Open Sans"),
19         legend.position = "none")

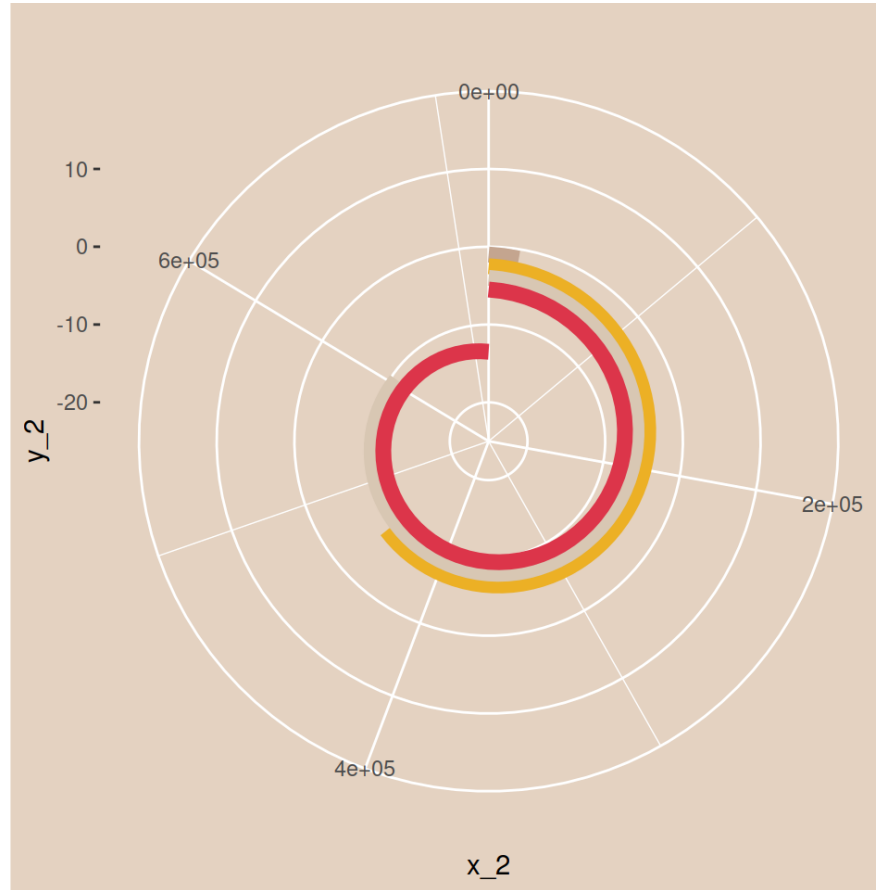
```

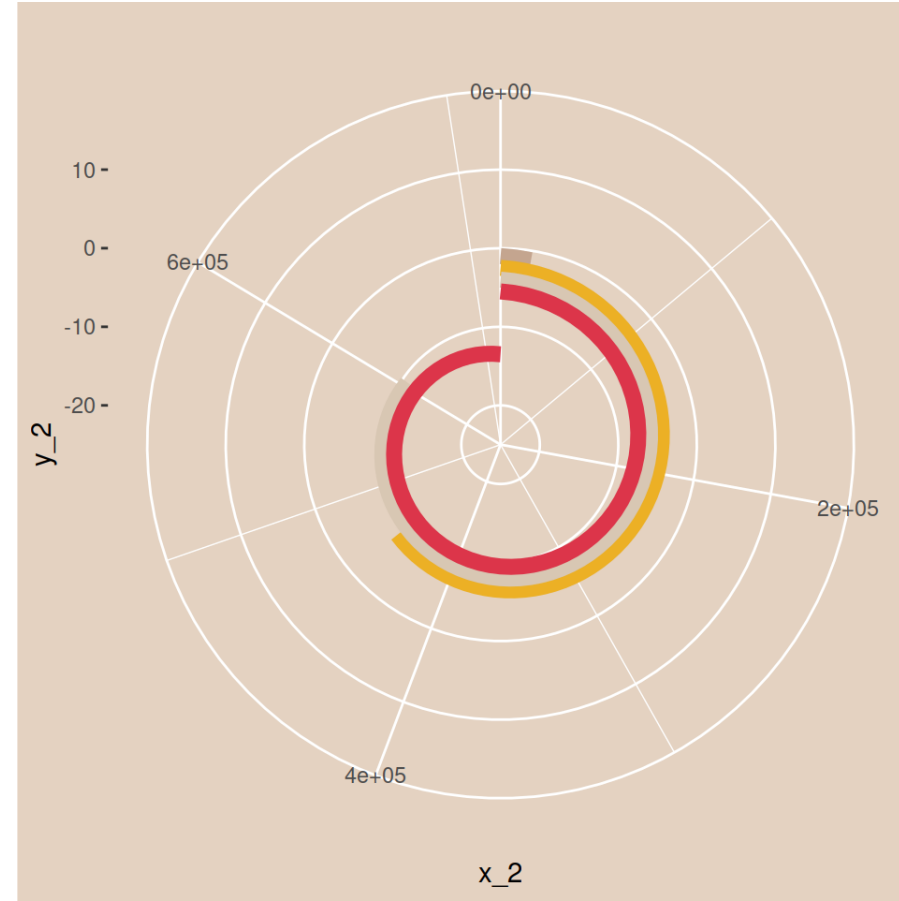
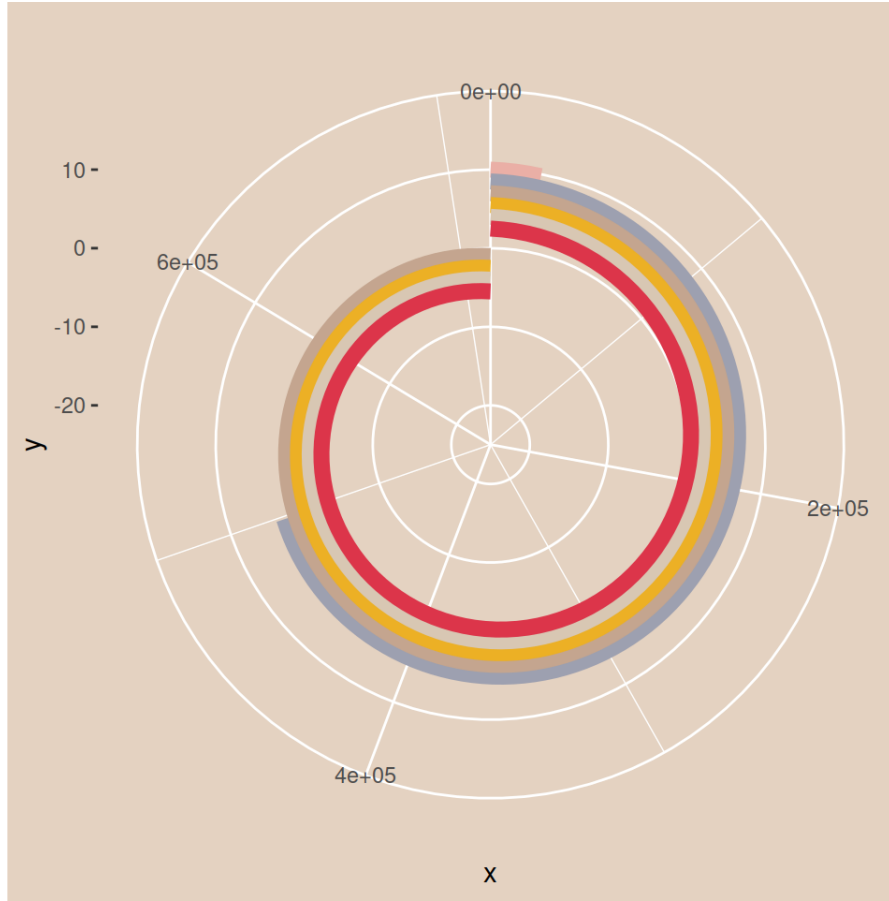



```

1  ggplot(data = furniture_spiral) +
2    # second part of spiral
3    geom_segment(aes(x = x_2, xend = xend_2,
4                    y = y_2, yend = yend_2,
5                    color = year),
6                size = 3) +
7    coord_polar(clip = "off") +
8    ylim(-25, 15) +
9    xlim(0, 717487.5) +
10   scale_color_manual(values = dubois_palette) +
11   theme(plot.background = element_rect(fill = bcgrnd,
12                                       color = NA),
13         plot.margin = margin(t = 20, r = 5, b = 5, l = 5),
14         panel.background = element_rect(fill = bcgrnd,
15                                         color = NA),
16         plot.title = element_text(hjust = 0.5, family = "Cutive",
17                                   face = "bold", size = 15),
18         plot.caption = element_text(size = 4, family = "Open Sans"),
19         legend.position = "none")

```





```

1 max_x = max(furniture$household_value_dollars) / 2
2
3 # Slope calculated using year 1885
4 slope = (-1 - 7) / (717487.5 - 0)
5
6 furniture_spiral = furniture %>%
7   mutate(year = as.factor(year),
8           y = seq(10, by = -1.5, length.out = 6),
9           x = 0) %>%
10  rowwise() %>%
11  mutate(xend = min(household_value_dollars, max_x),
12         yend = slope*xend + y) %>%
13  mutate(y_2 = yend,
14         x_2 = 0,
15         xend_2 = if_else(household_value_dollars < max_x,
16                          NA_real_, household_value_dollars - max_x),
17         yend_2 = slope*xend_2 + y_2)

```

```

# A tibble: 6 × 10
# Rowwise:
  year household_value_dollars      y      x      xend  yend  y_2  x_2  xend_2
  <fct>          <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1875           21186    10     0    21186    9.76    9.76     0     NA
2 1880           498532    8.5     0   498532    2.94    2.94     0     NA
3 1885           736170     7     0   717488.   -1     -1       0   18682.
4 1890           1173624    5.5     0   717488.  -2.5   -2.5     0  456136.
5 1895           1322694     4     0   717488.   -4     -4       0  605206.
6 1899           1434975    2.5     0   717488.  -5.5   -5.5     0  717488.
# i 1 more variable: yend_2 <dbl>

```

The y-axis is created by us

Especially for those segments that don't do a full spin (=they don't end at a nice "round" value like -1), we have to calculate the corresponding y-axis value for the x-axis value that we have

For this, we need slope


```

1 furniture_label = furniture_spiral %>%
2   select(year, household_value_dollars, y, x) %>%
3   mutate(dollars =
4     scales::dollar(
5       household_value_dollars,
6       prefix = "$"),
7     label =
8     if_else(
9       year %in% c("1880", "1885"),
10      glue("{year} ----- {dollars}"),
11      if_else(year == "1875",
12              glue("{year} ----- {dollars}"),
13              glue("{year} --- {dollars}")))

```

This is just some formatting of the text annotations on the plot

Let's combine all parts now!

```
1 ggplot(data = furniture_spiral) +
2   # first part of spiral
3   geom_segment(aes(x = x, xend = xend,
4                     y = y, yend = yend,
5                     color = year),
6                 size = 3) +
7   # second part of spiral
8   geom_segment(aes(x = x_2, xend = xend_2,
9                     y = y_2, yend = yend_2,
10                    color = year),
11                size = 3) +
12  geom_text(data = furniture_label,
13            aes(x = x, y = y,
14                label = label),
15              family = "Roboto Condensed",
16              size = 2.5,
17              hjust = 1) +
```

```

1  labs(title = paste0("ASSESSED VALUE OF HOUSEHOLD\n",
2                      "AND KITCHEN FURNITURE\n",
3                      "OWNED BY GEORGIA NEGROES.)) +
4  scale_color_manual(values = dubois_palette) +
5  coord_polar(clip = "off") +
6  ylim(-25, 15) +
7  xlim(0, 717487.5) +
8  theme_void() +
9  theme(plot.background = element_rect(fill = bcgrnd,
10                                         color = NA),
11        plot.margin = margin(t = 20, r = 5, b = 5, l = 5),
12        panel.background = element_rect(fill = bcgrnd,
13                                         color = NA),
14        plot.title = element_text(hjust = 0.5, family = "Cutive",
15                                   face = "bold", size = 15),
16        plot.caption = element_text(size = 4, family = "Open Sans"),
17        legend.position = "none"
18  )

```

ASSESSED VALUE OF HOUSEHOLD AND KITCHEN FURNITURE OWNED BY GEORGIA NEGROES.





What is Makeover Monday?

Join us in the weekly #MakeoverMonday series. Each week we post a link to a chart, and its data, and then you rework the chart. Maybe you retell the story more effectively, or find a new story in the data. We're curious to see the different approaches you all take. Whether it's a simple bar chart or an elaborate infographic, we encourage everyone of all skills to partake. Together we can have broader conversations about and with data.

<https://data.world/makeovermonday>

Until next week

Go back to the data set with more than 10 000 and less than 100 000 rows that you selected and prepared in different storage formats in the first units and load it

Use the decision tree at <https://www.data-to-viz.com/> to get inspired what kind of visualization you could do with it

Implement the visualization using packages in the programming language of your choice

Acknowledgements

https://twitter.com/ijeamaka_a/status/1361715338027560962

https://github.com/Ijeamakaanyene/tidyuesday/blob/master/scripts/2021_06_dubois_data.Rmd

<https://github.com/charlie-gallagher/tidy-tuesday/tree/master>