

Exercise 8 | Grammar of Graphics I

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

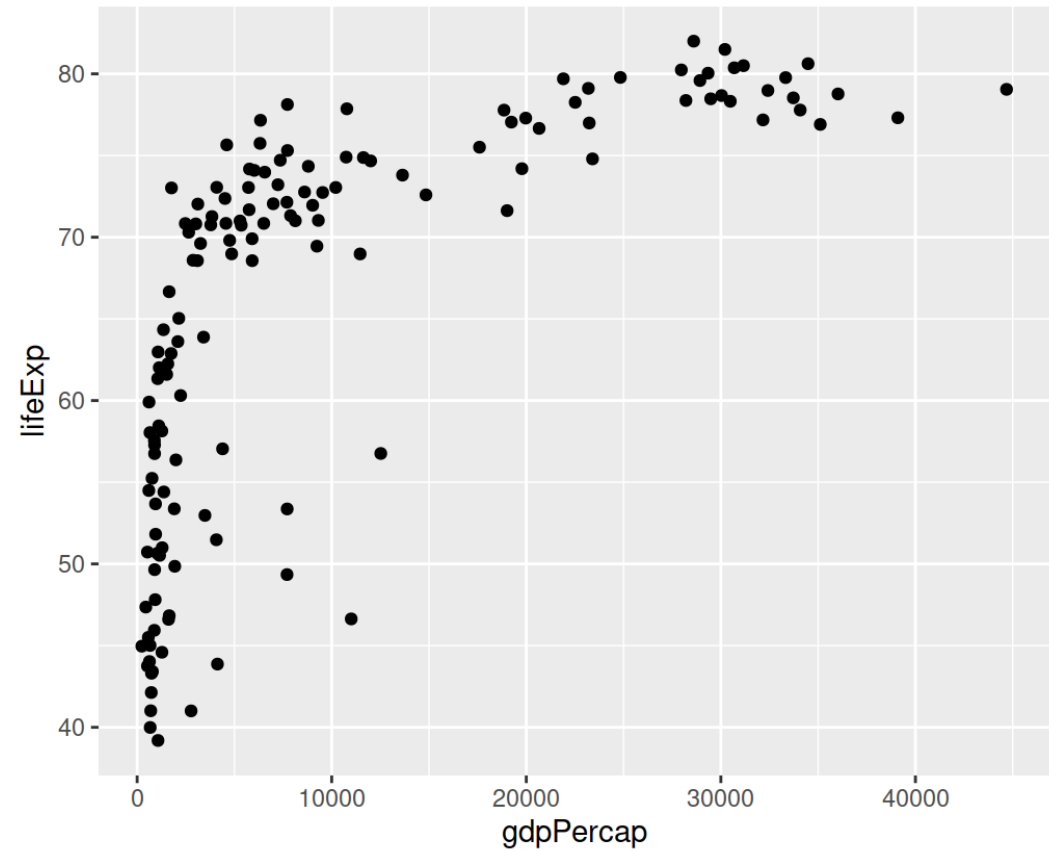
IS 616: Large Scale Data Analysis and Visualization

For the following, we consider one year only from the `gapminder` data:

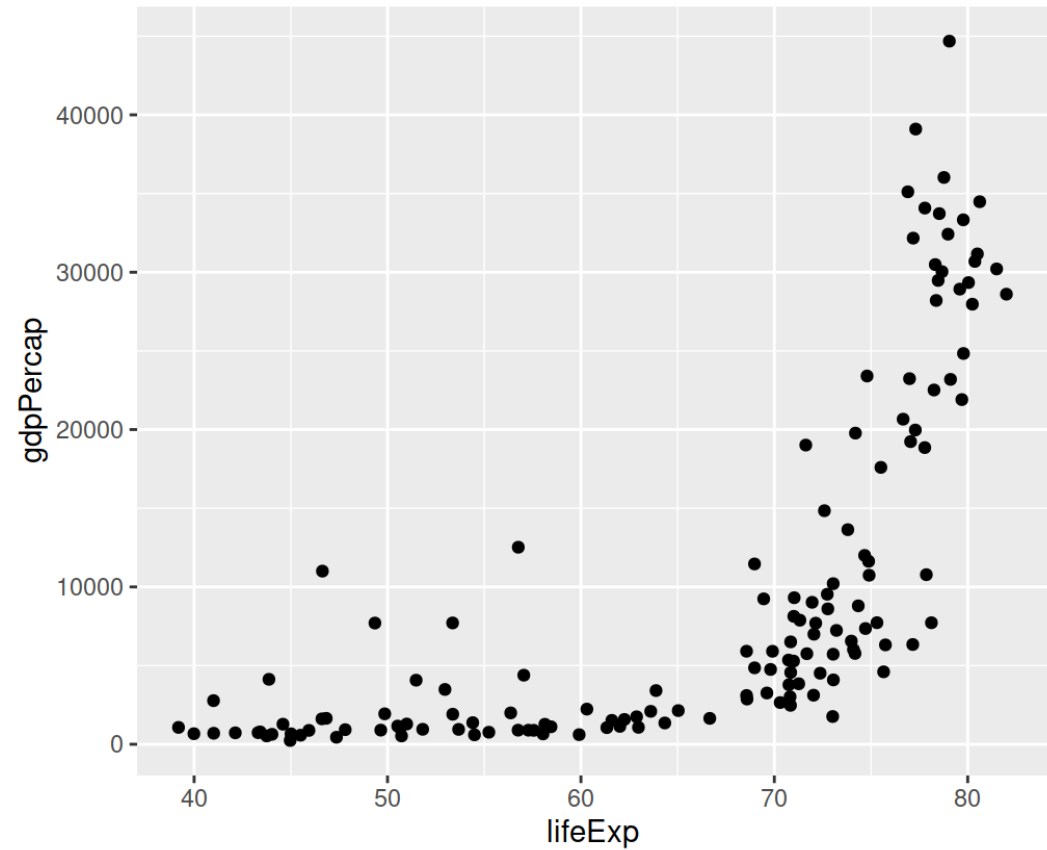
```
1 gapminder_2002 <- gapminder[gapminder$year == 2002, ]
```

Let's use this subset of the data set to explore different functions in `ggplot2` that modify the coordinate system

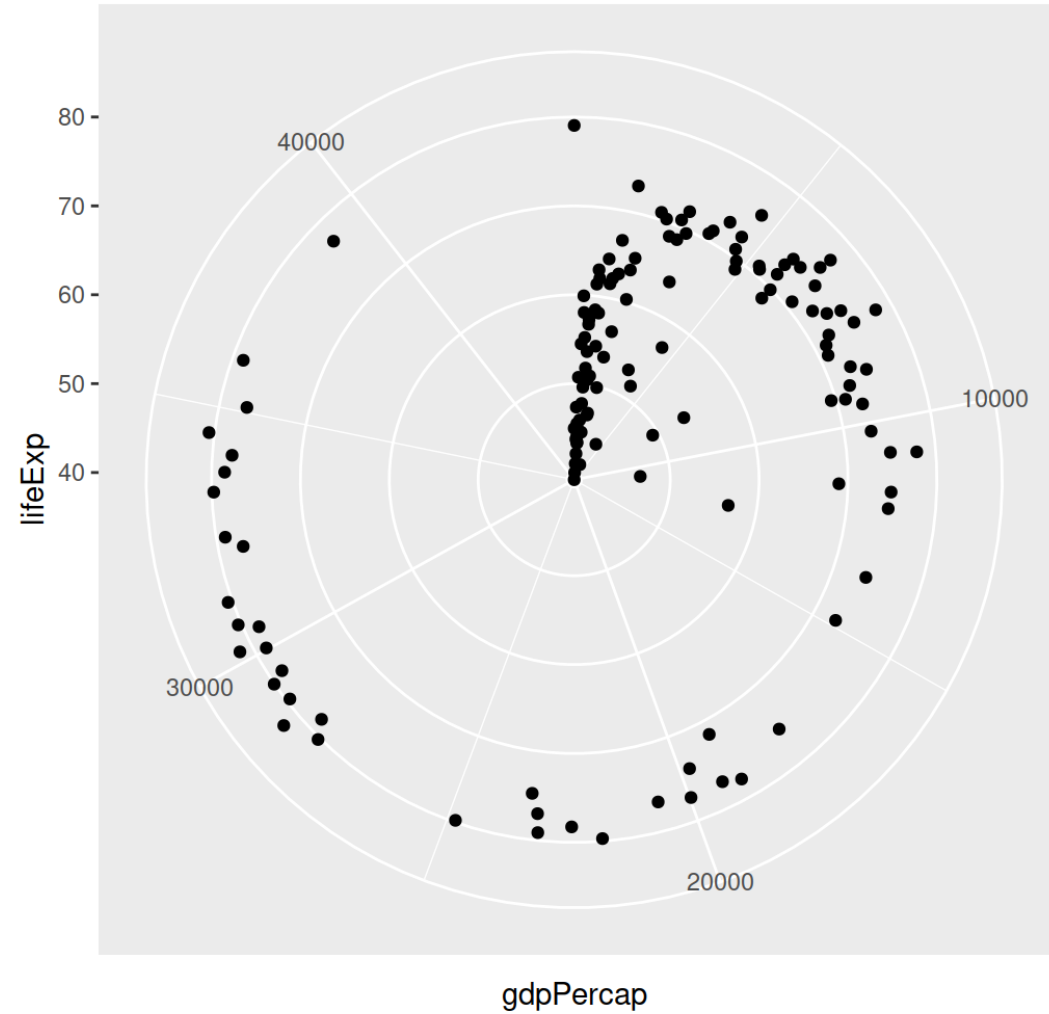
```
1 ggplot(data =
2       gapminder_2002) +
3   aes(x = gdpPercap,
4       y = lifeExp) +
5   geom_point() +
6   coord_cartesian()
```



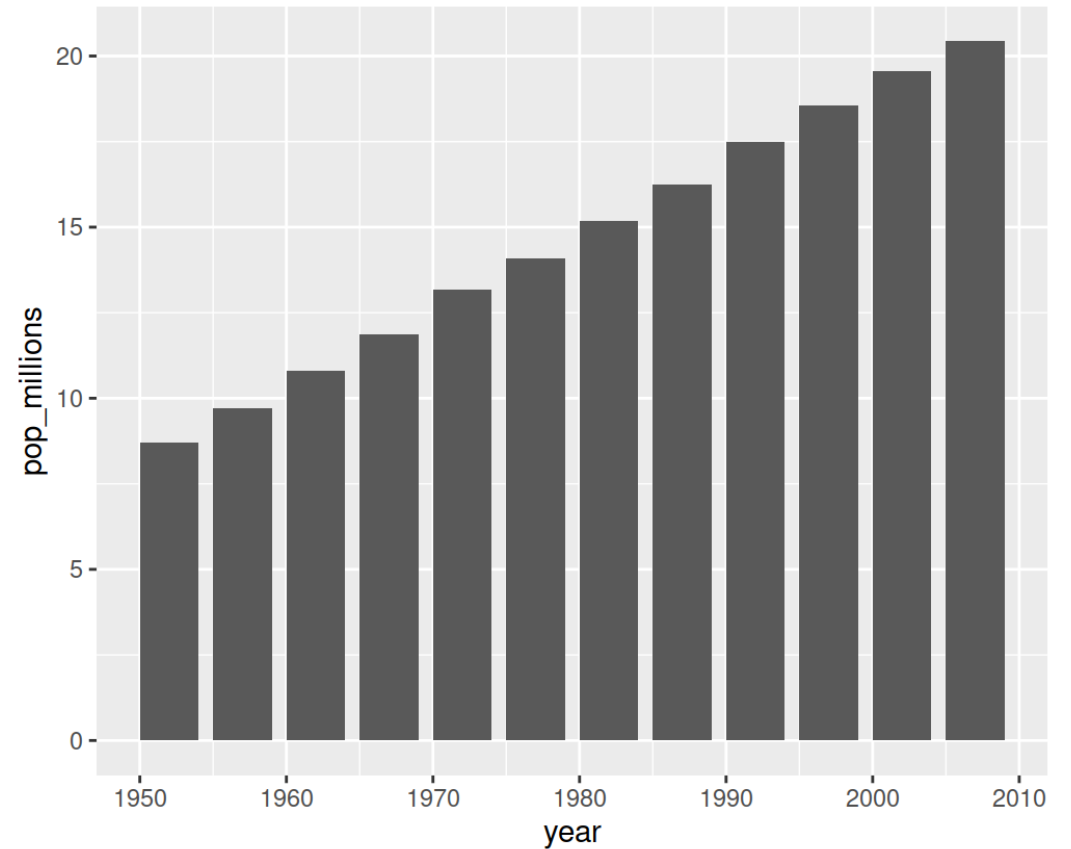
```
1 ggplot(data =
2       gapminder_2002) +
3   aes(x = gdpPercap,
4       y = lifeExp) +
5   geom_point() +
6   coord_flip()
```



```
1 ggplot(data =
2       gapminder_2002) +
3   aes(x = gdpPercap,
4       y = lifeExp) +
5   geom_point() +
6   coord_polar()
```



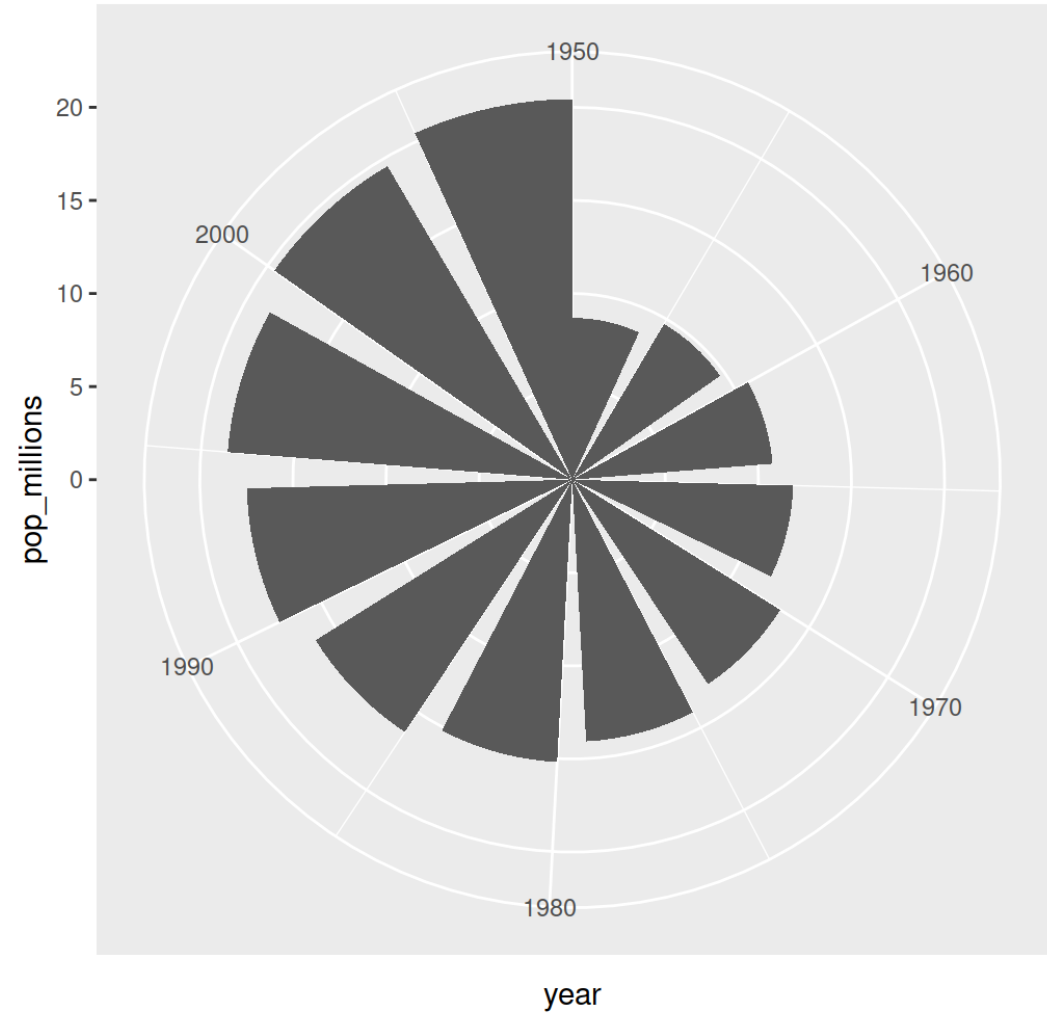
```
1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5         = pop/1000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10     width = 4,
11     position = "dodge")
```



```

1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5          = pop/10000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10      width = 4,
11      position = "dodge") +
12  coord_polar()

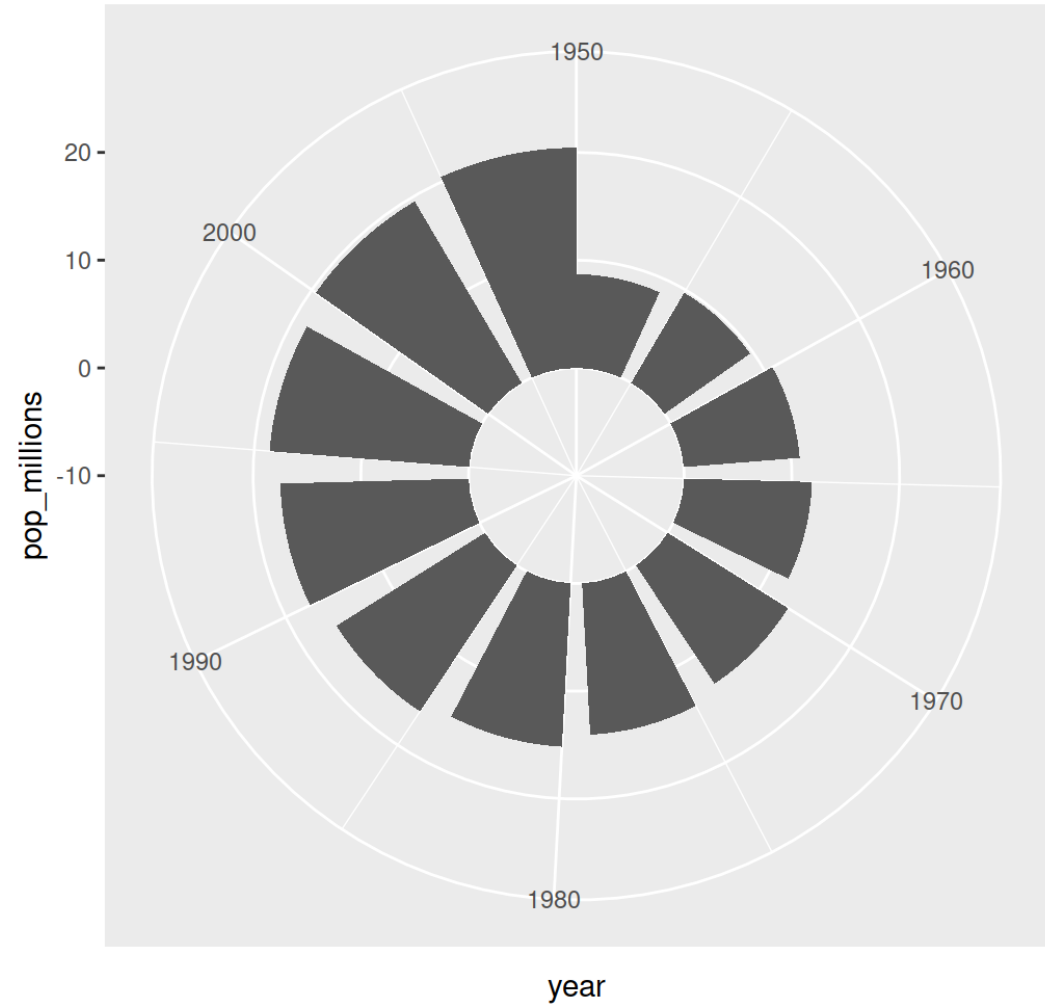
```



```

1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5          = pop/1000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10     width = 4,
11     position = "dodge") +
12   coord_polar() +
13   scale_y_continuous(
14     limits = c(-10, 25))

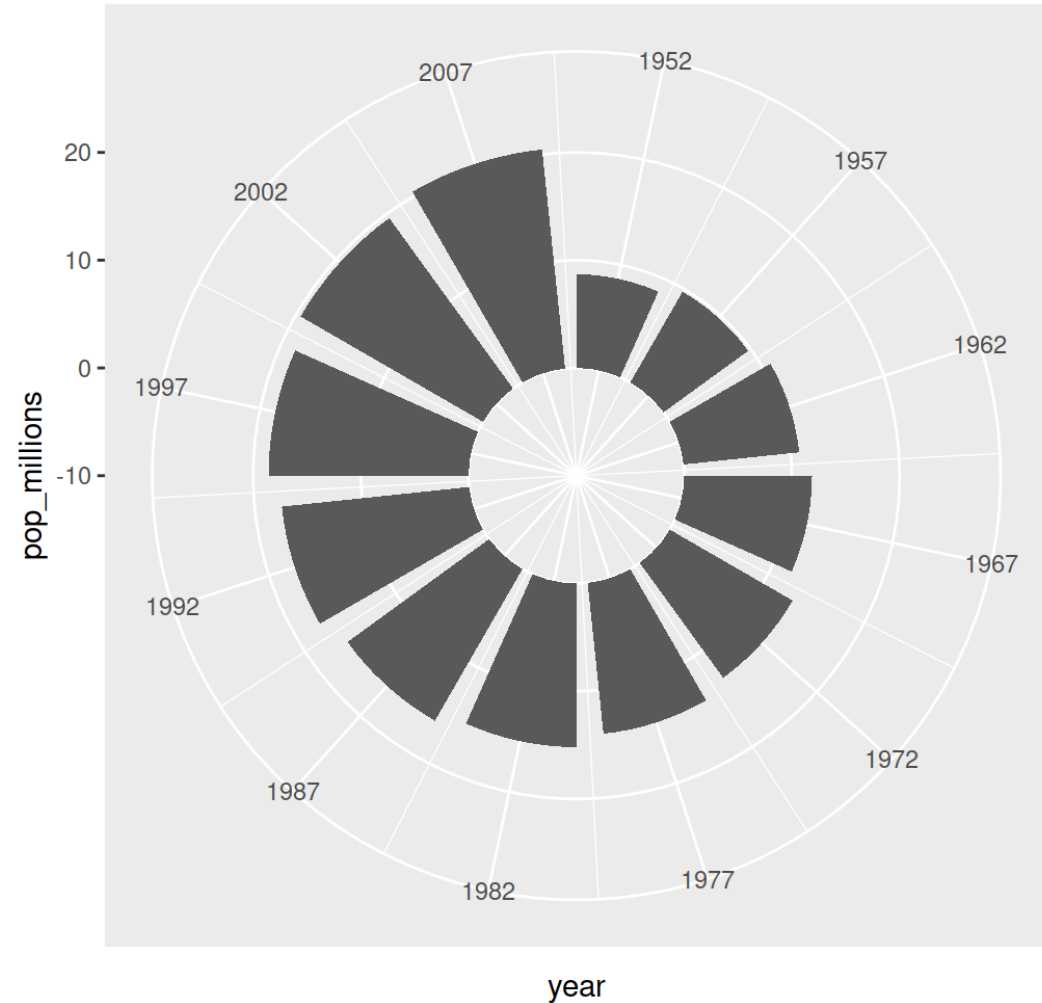
```




```

1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5         = pop/1000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10    width = 4,
11    position = "dodge") +
12  coord_polar() +
13  scale_y_continuous(
14    limits = c(-10, 25)) +
15  scale_x_continuous(
16    breaks =
17      seq(1952, 2007, by = 5),
18    limits =
19      c(1950, 2010))

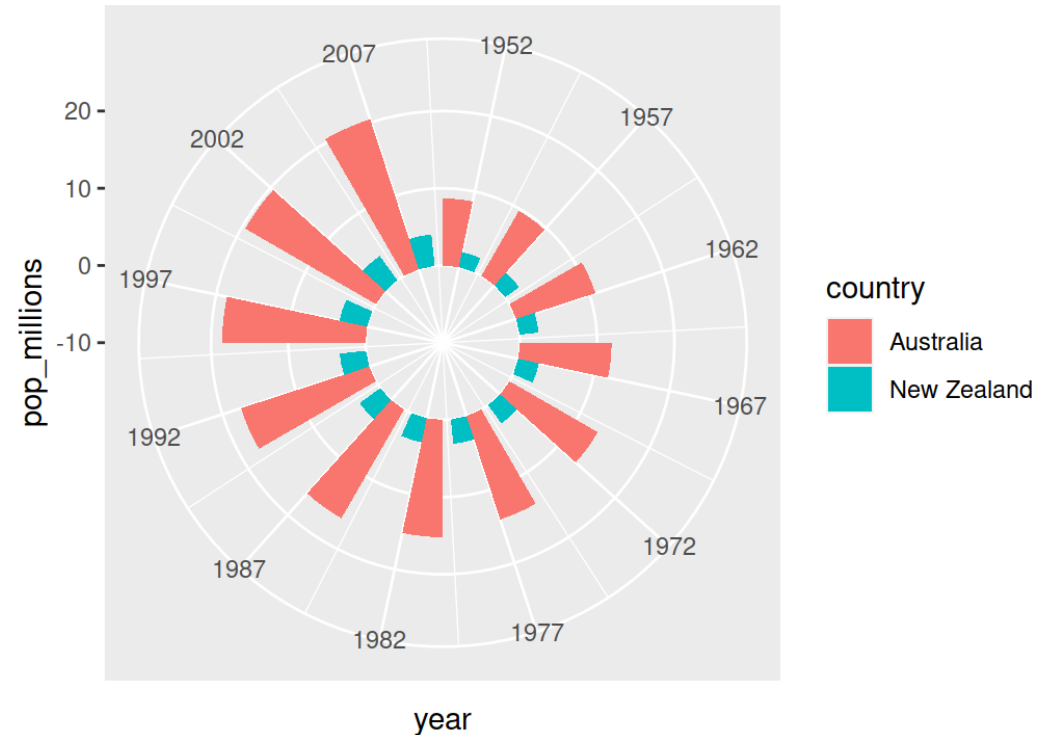
```



```

1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5         = pop/1000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10    width = 4,
11    position = "dodge") +
12  coord_polar() +
13  scale_y_continuous(
14    limits = c(-10, 25)) +
15  scale_x_continuous(
16    breaks =
17      seq(1952, 2007, by = 5),
18    limits =
19      c(1950, 2010)) +
20  aes(fill = country)

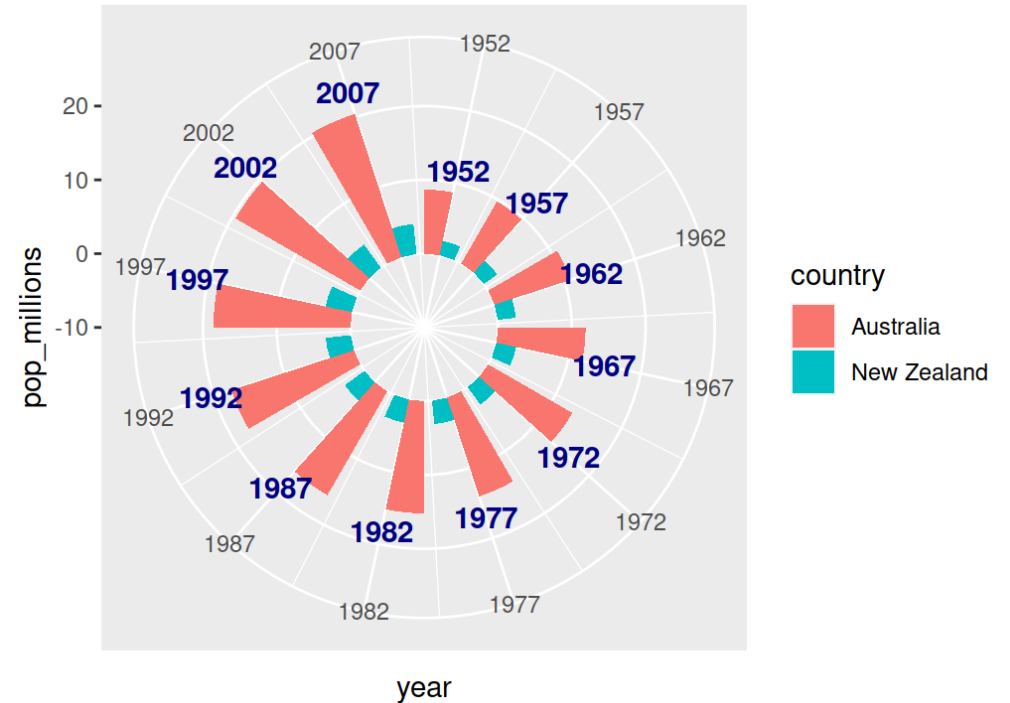
```



```

1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5         = pop/10000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10    width = 4,
11    position = "dodge") +
12  coord_polar() +
13  scale_y_continuous(
14    limits = c(-10, 25)) +
15  scale_x_continuous(
16    breaks =
17      seq(1952, 2007, by = 5),
18    limits =
19      c(1950, 2010)) +
20  aes(fill = country) +

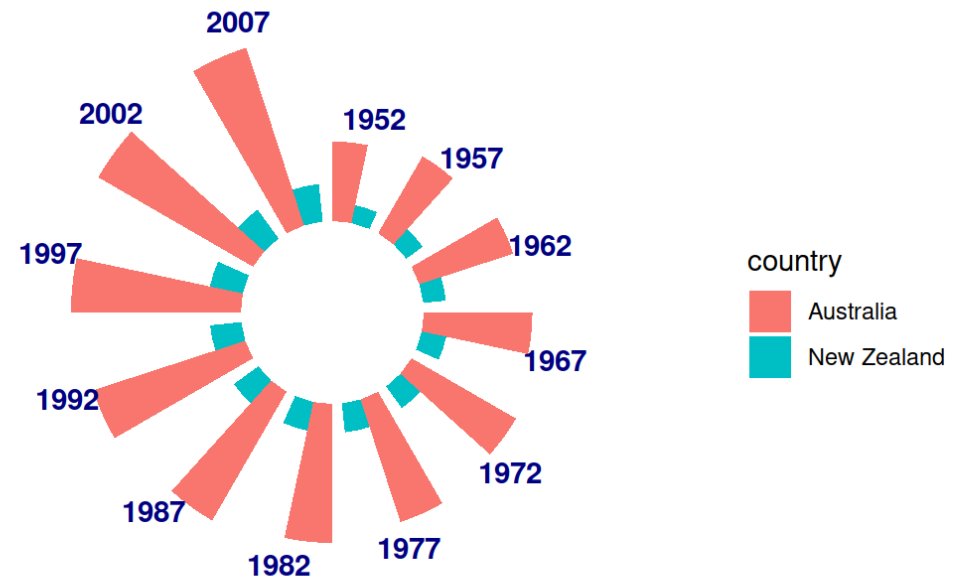
```



```

1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5         = pop/10000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10    width = 4,
11    position = "dodge") +
12  coord_polar() +
13  scale_y_continuous(
14    limits = c(-10, 25)) +
15  scale_x_continuous(
16    breaks =
17      seq(1952, 2007, by = 5),
18    limits =
19      c(1950, 2010)) +
20  aes(fill = country) +

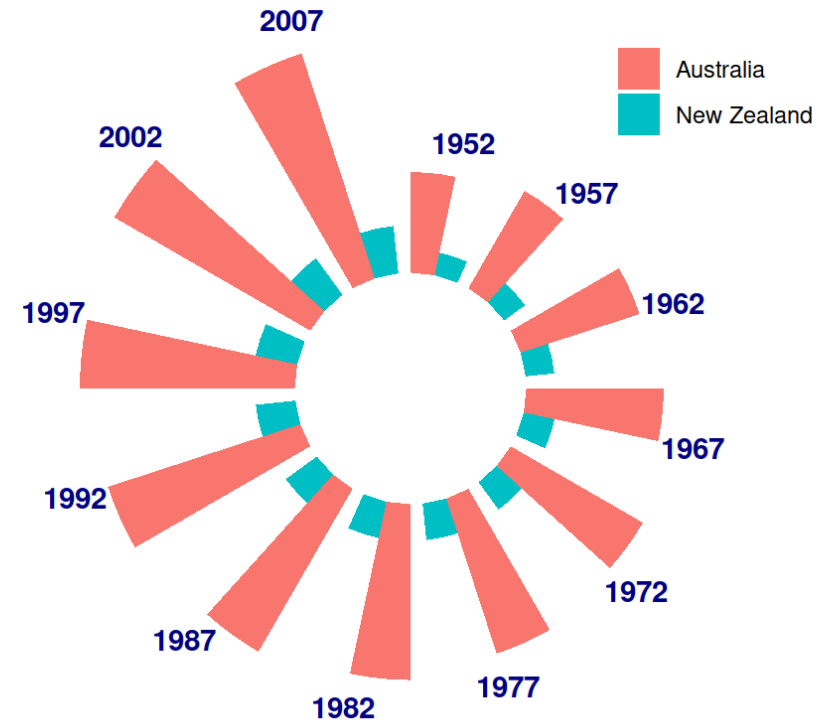
```

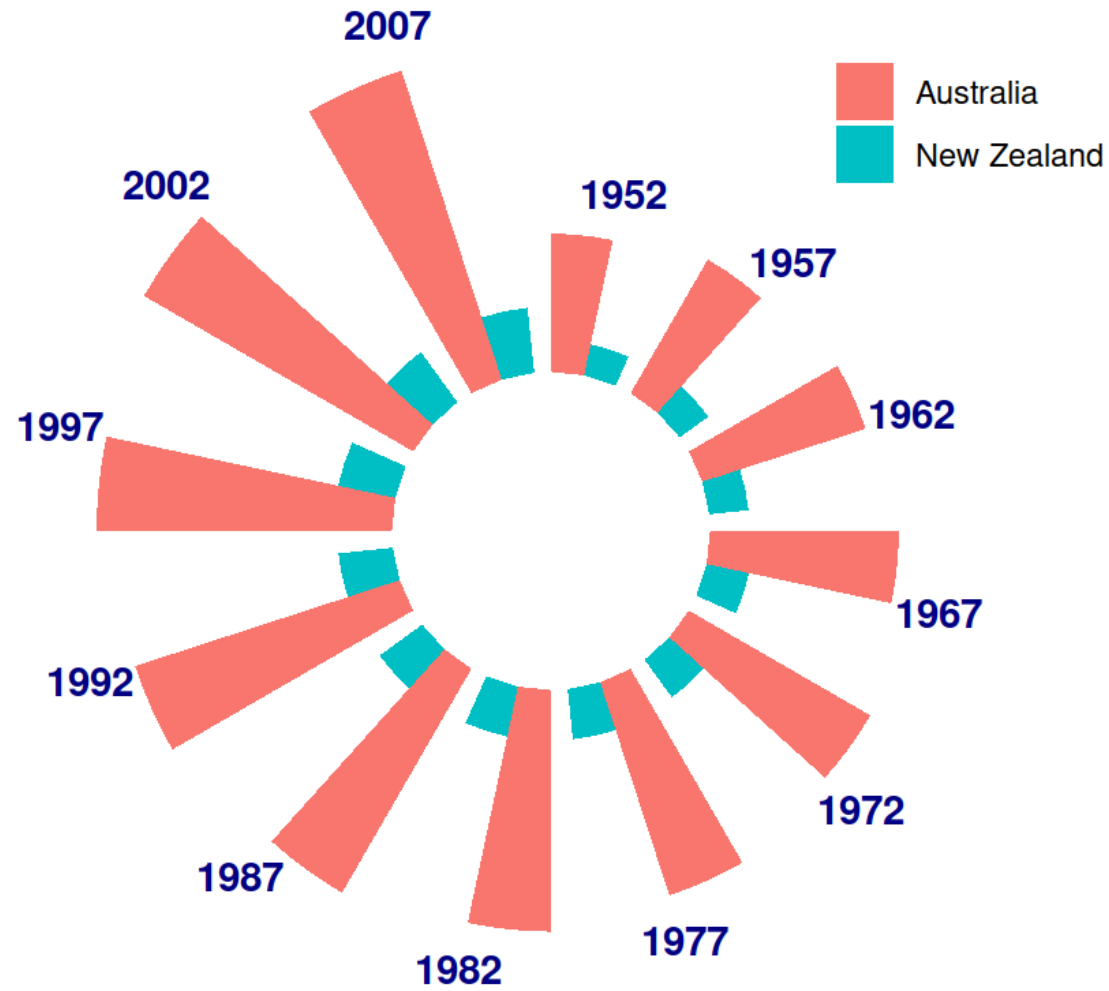


```

1 gapminder %>%
2   filter(continent
3         == "Oceania") %>%
4   mutate(pop_millions
5         = pop/10000000) %>%
6   ggplot() +
7   aes(x = year) +
8   aes(y = pop_millions) +
9   geom_col(
10     width = 4,
11     position = "dodge") +
12   coord_polar() +
13   scale_y_continuous(
14     limits = c(-10, 25)) +
15   scale_x_continuous(
16     breaks =
17       seq(1952, 2007, by = 5),
18     limits =
19       c(1950, 2010)) +
20   aes(fill = country) +

```





About last week's exercise

https://en.wikipedia.org/wiki/Simpson%27s_paradox



▶ 0:00 / 3:36




```
1 b <- fread("https://waf.cs.illinois.edu/discovery/berkeley.csv")
2 gender_admission_table <- table(b[,.(Gender, Admission)])
3 kable(gender_admission_table)
```

	Accepted	Rejected
F	1494	2827
M	3738	4704

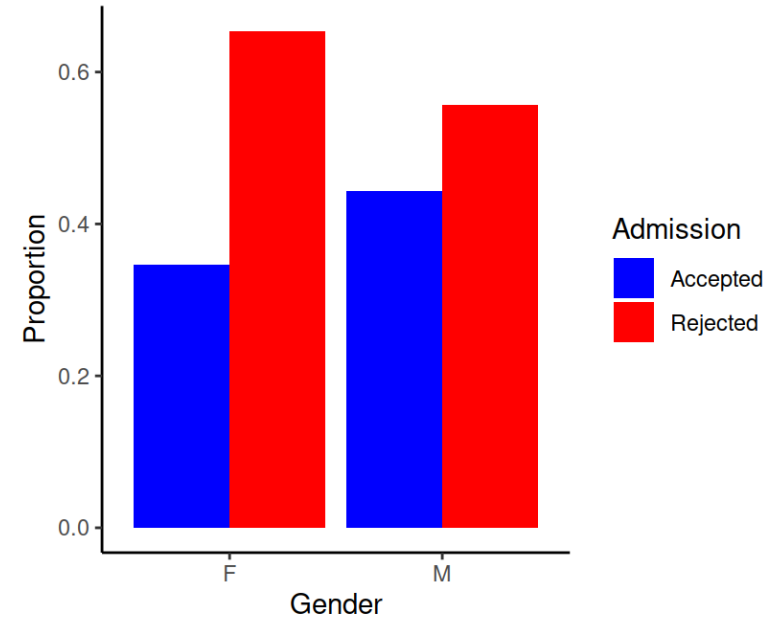
```
1 kable(round(proportions(gender_admission_table,margin=1),2))
```

	Accepted	Rejected
F	0.35	0.65
M	0.44	0.56

```

1 total_app_gender <-
2   b[,.(total_N=.N),Gender]
3
4 b_adm_gender <- b[,.N,by=
5   c("Gender","Admission")]
6
7 b_prop_adm_gender <- merge(
8   b_adm_gender, total_app_gender,
9   by="Gender")[,.(Gender,Admission,
10    Proportion=N/total_N)]
11
12 ggplot(b_prop_adm_gender,
13   aes(x=Gender,y=Proportion,
14     fill=Admission)) +
15   geom_col(position = "dodge") +
16   scale_fill_manual(
17     values=c("blue","red")) +
18   theme_classic()

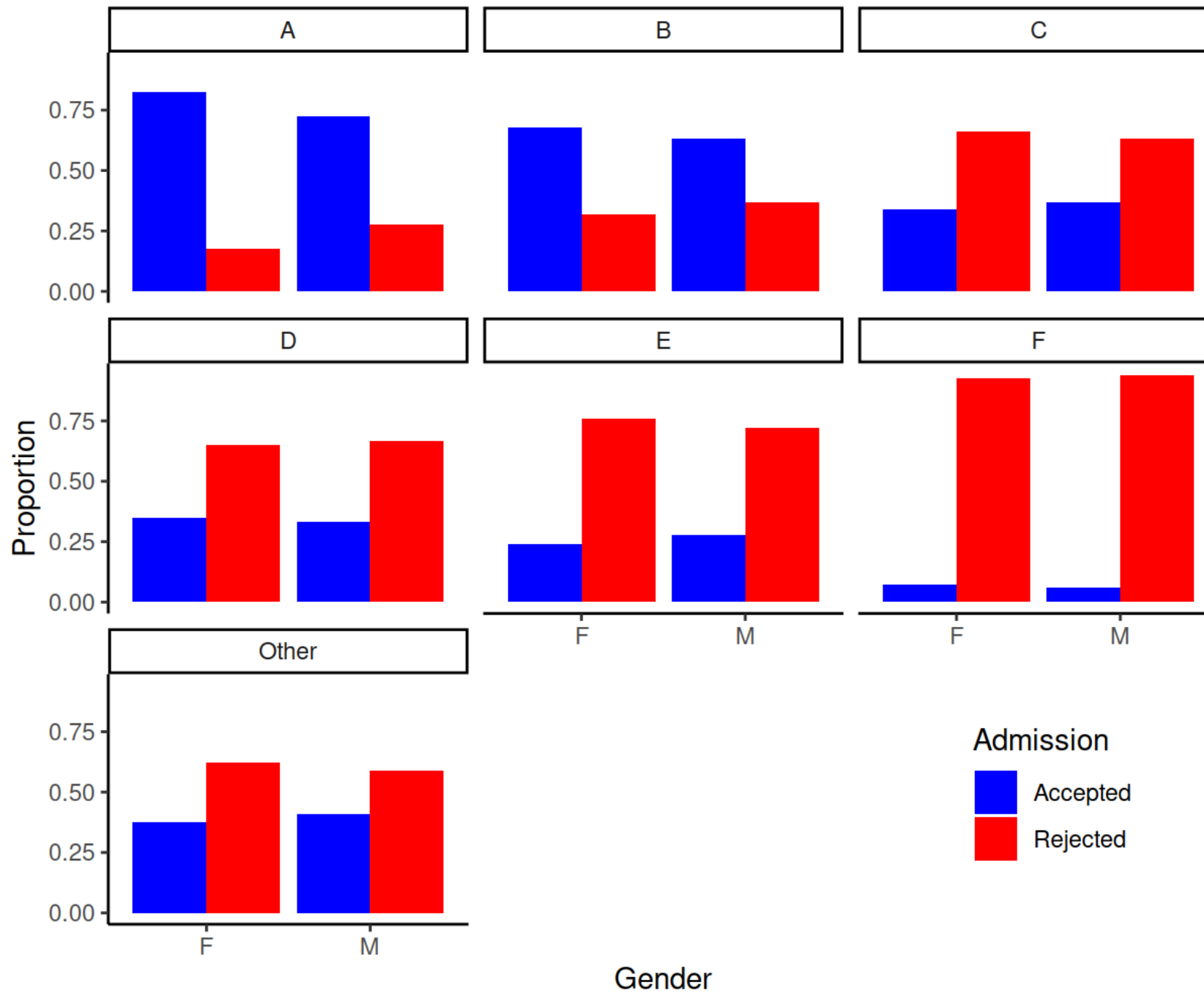
```



```

1 total_app_gender_major <-
2   b[,.(total_N=.N),c("Major","Gender")]
3
4 b_adm_gender_major <- b[,.N,by=
5   c("Major","Gender","Admission")]
6
7 b_prop_adm_gender_major <- merge(
8   b_adm_gender_major, total_app_gender_major,
9   by=c("Major","Gender"))[,.(Major,Gender,Admission,
10  Proportion=N/total_N)]
11
12 ggplot(b_prop_adm_gender_major,
13        aes(x=Gender, y=Proportion, fill=Admission)) +
14  geom_col(position = "dodge") +
15  scale_fill_manual(
16    values=c("blue","red")) +
17  theme_classic() +
18  facet_wrap(~ Major) +
19  theme(legend.position = c(0.85, 0.15))

```



```

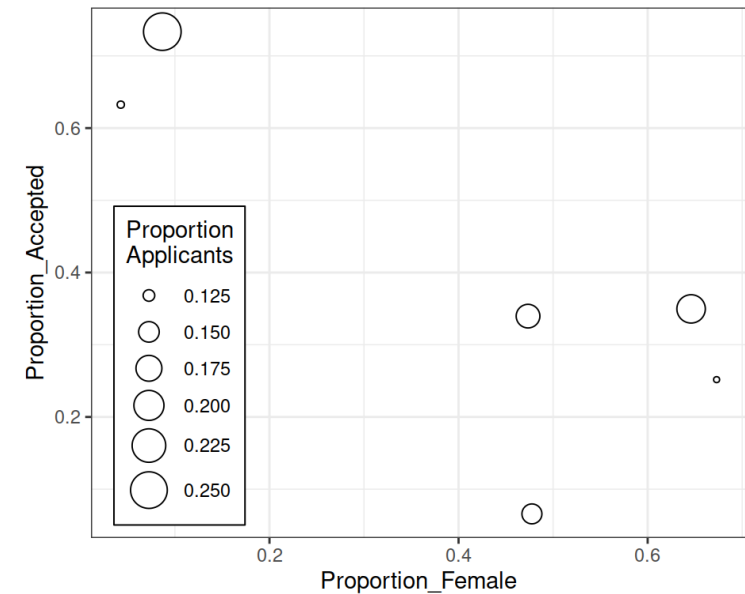
1 b <- b[Major!="Other"]
2
3 total_app_gender_major <-
4   b[,.(N_gender=.N),,c("Major","Gender")]
5 total_major <- b[,.(total_N_major=.N),c("Major")]
6 prop_female_major <-
7   merge(total_app_gender_major,
8         total_major,by="Major")[Gender=="F",
9         .(Proportion_Female=N_gender/total_N_major),Major]
10
11 admission_major <-
12   data.table(proportions(table(b[,.(Admission, Major)]),margin = 2)
13             )[Admission == "Accepted",.(Proportion_Accepted=N,Major)]
14
15 prop_app_major <- data.table(proportions(table(b[,.(Major)])))
16 colnames(prop_app_major) <- c("Major","Proportion_Applicants")
17
18 pldt_major <- merge(merge(admission_major,prop_female_major,
19                          by="Major"),prop_app_major,by="Major")

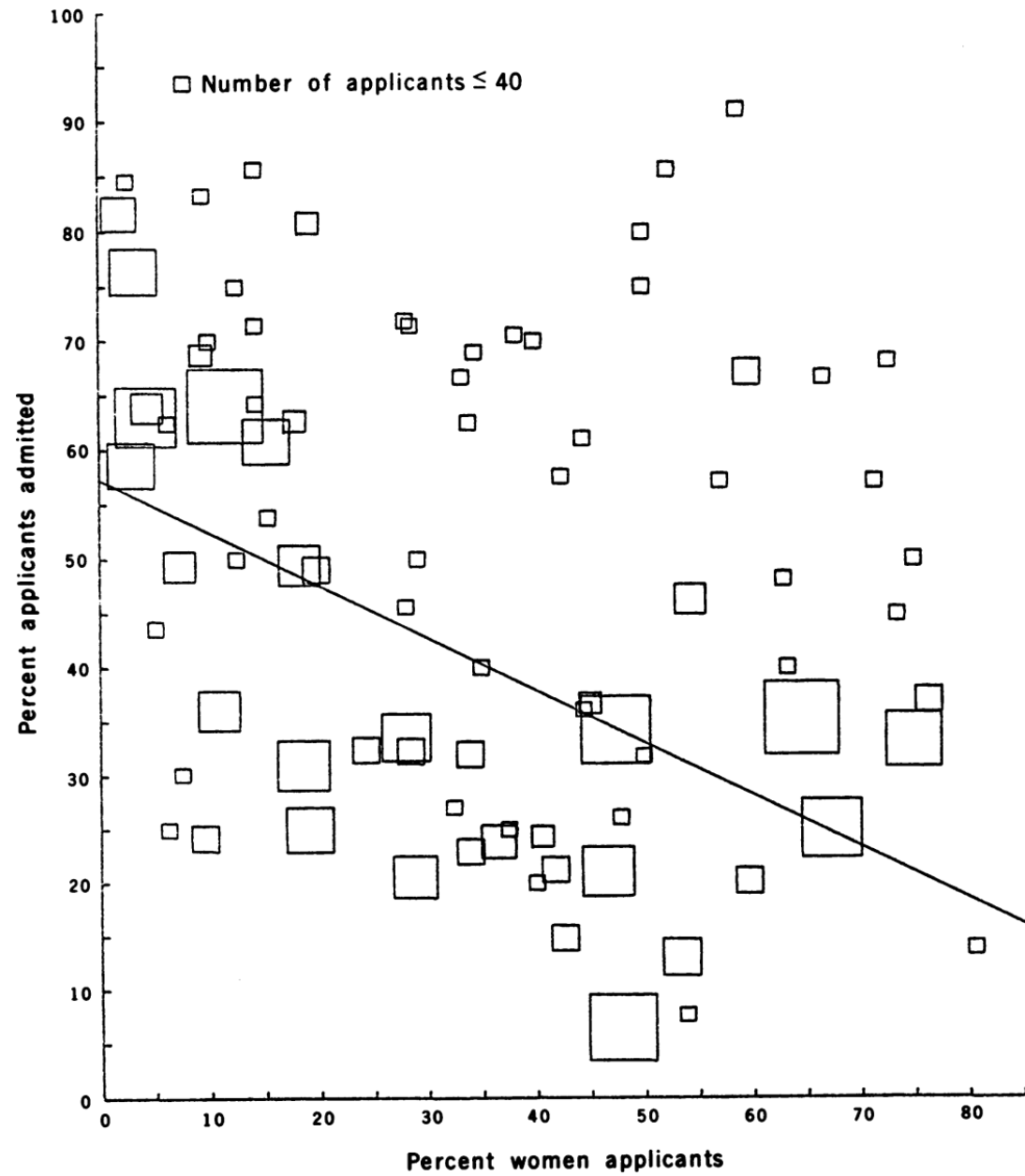
```

```

1  ggplot(pldt_major,
2      aes(x=Proportion_Female,
3          y=Proportion_Accepted,
4          size=Proportion_Applicants)
5      ) +
6  geom_point(fill="white",
7            color="black",
8            shape = 21) +
9  theme_bw() +
10 guides(
11     size=guide_legend(
12         "Proportion\nApplicants")) +
13 theme(
14     legend.position = c(0.135, 0.325),
15     legend.background =
16         element_rect(
17             colour = 'black',
18             linewidth = 0.35)) +
19 scale_size(range = c(1, 8))

```





“Proportion of applicants that are women plotted against proportion of applicants admitted, in 85 departments. Size of box indicates relative number of applicants to the department.”

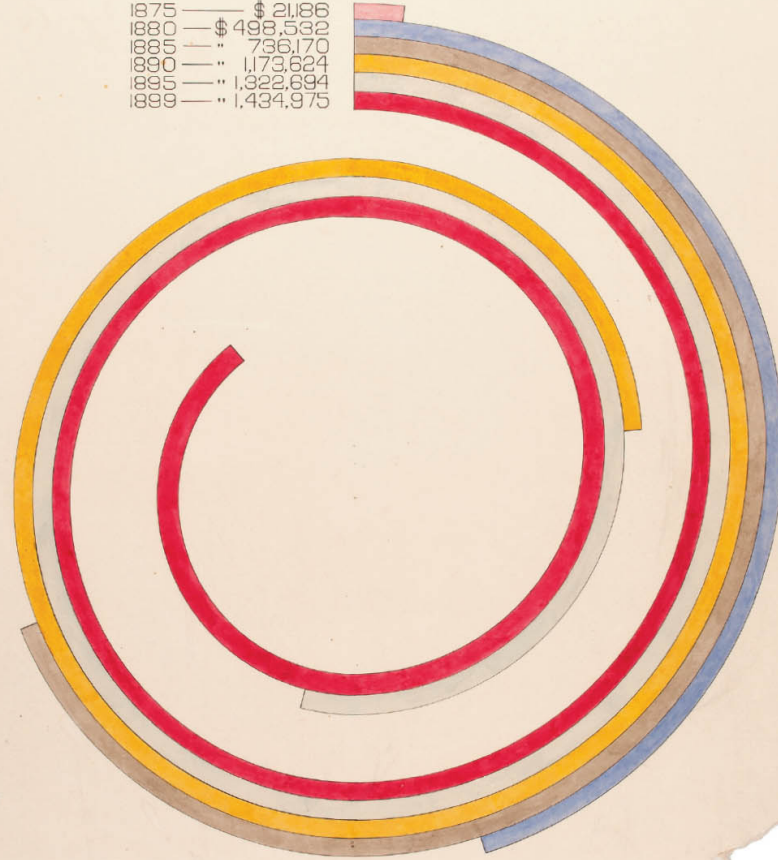
Bickel, P. J., Hammel, E. A., & O’Connell, J. W. (1975). Sex Bias in Graduate Admissions: Data from Berkeley: Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation. *Science*, 187(4175), 398–404.

<https://doi.org/10.1126/science.187.4175.398>

Until next week...

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

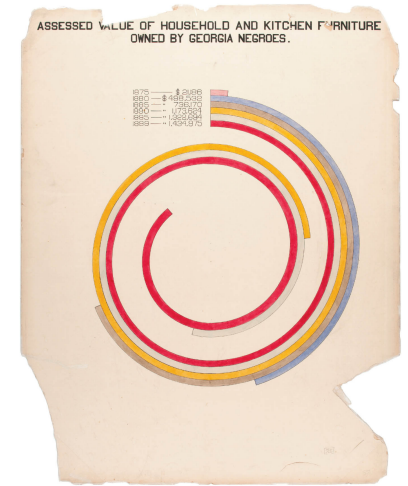
1875	—	\$ 21,886
1880	—	\$ 498,532
1885	—	736,170
1890	—	1,173,624
1895	—	1,322,694
1899	—	1,434,975



“Six curving bars of color initiate from a stacked block of text listing the value of household furniture owned by black Georgians over a twenty-five-year period.

Each spiral is a different pastel or primary color that wraps around a core origin point.

The rings show a growing trend over time, each longer than the previous one.



The unusual and complex configuration of the spiral diagram here builds on graphic constructions such as Playfair’s pie chart and Nightingale’s rose diagram. The end result is simultaneously easy to read and hypnotic.”

Du Bois, W. E. B., Battle-Baptiste, W., & Rusert, B. (2018). W.E.B. Du Bois’s data portraits: Visualizing Black America the color line at the turn of the twentieth century (First edition). The W.E.B. Du Bois Center at the University of Massachusetts Amherst Princeton Architectural press.

Replicate “Assessed Value of Household and Kitchen Furniture”
(Plate 25)

Hint: `coord_polar` is your friend

Data was provided as part of efforts to replicate all plates:

<https://github.com/ajstarks/dubois-data-portraits/tree/master>

See this article for more context:

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

furniture.csv

variable	class	description
Year	double	Year
Houshold Value (Dollars)	double	Furniture value

<https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2021/2021-02-16/furniture.csv>

Acknowledgements

https://evamaerey.github.io/ggplot2_grammar_guide/about