Case Study III

Alex Christmann

2022-07-14

Question 1

Nightingale's rose chart was remarkably innovative for its time (1859). What makes it special is how the circular design lets you easily compare different time periods at a glance. The clever use of adjacent colors helps anyone quickly see the proportion of deaths from different causes. Following the chart from April 1854 to March 1856, you can clearly see the story of declining mortality as improvements were made. While this visualization packs in a lot of information beautifully, it does present some challenges for viewers who aren't used to this kind of chart. The split into two separate roses requires the viewer to mentally connect the time sequence across both charts. Despite its visual appeal and the rich story it tells, a well-designed bar chart might be easier for most audiences to understand while still showing the important patterns that drove Nightingale's key insights.

Question 2

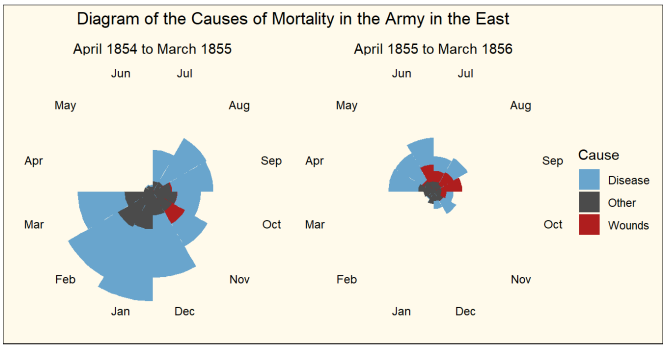
Below I've recreated Nightingale's famous rose chart using modern data tools. I've transformed the original dataset to show mortality patterns by cause from April 1854 through March 1856. This version stays true to Nightingale's original design while making it easier to analyze the patterns and see how different causes of death changed over time.

```
setwd("C:/Users/alexc/OneDrive/Documents/Summer 2022/Descriptive Analytics/R Markdowns")
load("~/Summer 2022/Descriptive Analytics/R Markdowns/Nightingale.RData")

library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.1
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.3.6     ✓ purrr   0.3.4
## ✓ tibble  3.1.7     ✓ dplyr   1.0.9
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
```

```
Nightingale %>%
  select(Date, Month, Year, contains("rate")) %>%
  pivot_longer(cols = 4:6, names_to = "Cause", values_to = "Rate") %>%
  mutate(Cause = gsub(".rate", "", Cause),
         period = ifelse(Date <= as.Date("1855-03-01"), "April 1854 to March 1855", "April 1855 to March 1856"),
         Month = fct_relevel(Month, "Jul", "Aug", "Sep", "Oct", "Nov", "Dec", "Jan", "Feb", "Mar", "Apr", "May", "Jun")) %>%
  ggplot(aes(Month, Rate)) +
  geom_col(aes(fill = Cause), width = 1, position = "identity") +
  coord_polar() +
  facet_wrap(~period) +
  scale_fill_manual(values = c("skyblue3", "grey30", "firebrick")) +
  scale_y_sqrt() +
  theme_void() +
  theme(axis.text.x = element_text(size = 9),
        strip.text = element_text(size = 11),
        legend.position = "right",
        plot.background = element_rect(fill = alpha("cornsilk", 0.5)),
        plot.margin = unit(c(10, 10, 10, 10), "pt"),
        plot.title = element_text(vjust = 5)) +
  ggtitle("Diagram of the Causes of Mortality in the Army in the East") +
  theme(plot.title = element_text(hjust = 0.5))
```



Looking at this data, we can see a dramatic drop in deaths that lines up with when Nightingale implemented her sanitation improvements. However, as a data scientist, I need to point out that correlation doesn't always mean causation. The fact that deaths were already starting to decline before her changes suggests other factors might have been at work too. While the size of the improvement strongly suggests her methods were effective, it would be oversimplifying to give her reforms 100% of the credit without considering other possible influences like seasonal changes, other medical improvements, or changes in reporting. That said, the sharp reduction in disease-related deaths provides compelling evidence that her sanitation protocols made a substantial difference, even if we can't precisely measure exactly how much of the improvement was due to her work alone.

Question 3

I've created a more straightforward bar chart visualization below that makes the same data easier to understand. While Nightingale's rose chart was groundbreaking, this format allows anyone to immediately see the changes in mortality over time without having to mentally adjust for the circular format. The straightforward timeline helps tell the story of these mortality trends more directly.

```
library(tidyverse)
library(knitr)
library(grid)
library(HistData)
```

```
## Warning: package 'HistData' was built under R version 4.2.1
```

```
library(ggplot2)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

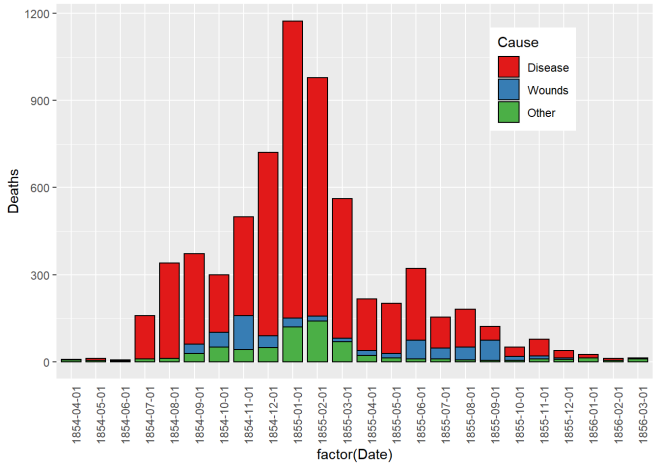
```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
rates <- Nightingale %>%
  as_tibble %>%
  subset(select = c(1, 8:10))

rates %<>%
  gather(key = "Cause", value = "Deaths", -Date) %>%
  mutate(Month = gl(12, 1, 72, labels = month.name[c(4:12, 1:3)])) %>%
  mutate(Practice = gl(2, 12, 72, labels = c("Old", "New"), ordered = TRUE))

rates$Cause %<>%
  sub("\\.rate", "", .)
rates$Cause %<>%
  factor(levels = c("Disease", "Wounds", "Other"))
rates %>%
  kable(align = c("c", "c", "r", "c", "c"))
```

Date	Cause	Deaths	Month	Practice
1854-04-01	Disease	1.4	April	Old
1854-05-01	Disease	6.2	May	Old
1854-06-01	Disease	4.7	June	Old
1854-07-01	Disease	150.0	July	Old
1854-08-01	Disease	328.5	August	Old
1854-09-01	Disease	312.2	September	Old
1854-10-01	Disease	197.0	October	Old
1854-11-01	Disease	340.6	November	Old
1854-12-01	Disease	631.5	December	Old
1855-01-01	Disease	1022.8	January	Old
1855-02-01	Disease	822.8	February	Old
1855-03-01	Disease	480.3	March	Old
1855-04-01	Disease	177.5	April	New
1855-05-01	Disease	171.8	May	New
1855-06-01	Disease	247.6	June	New
1855-07-01	Disease	107.5	July	New
1855-08-01	Disease	129.9	August	New
1855-09-01	Disease	47.5	September	New
1855-10-01	Disease	32.8	October	New
1855-11-01	Disease	56.4	November	New
1855-12-01	Disease	25.3	December	New
1856-01-01	Disease	11.4	January	New
1856-02-01	Disease	6.6	February	New
1856-03-01	Disease	3.9	March	New
1854-04-01	Wounds	0.0	April	Old
1854-05-01	Wounds	0.0	May	Old
1854-06-01	Wounds	0.0	June	Old
1854-07-01	Wounds	0.0	July	Old
1854-08-01	Wounds	0.4	August	Old
1854-09-01	Wounds	32.1	September	Old
1854-10-01	Wounds	51.7	October	Old
1854-11-01	Wounds	115.8	November	Old
1854-12-01	Wounds	41.7	December	Old
1855-01-01	Wounds	30.7	January	Old
1855-02-01	Wounds	16.3	February	Old
1855-03-01	Wounds	12.8	March	Old
1855-04-01	Wounds	17.9	April	New
1855-05-01	Wounds	16.6	May	New

Date	Cause	Deaths	Month	Practice
1855-06-01	Wounds	64.5	June	New
1855-07-01	Wounds	37.7	July	New
1855-08-01	Wounds	44.1	August	New
1855-09-01	Wounds	69.4	September	New
1855-10-01	Wounds	13.6	October	New
1855-11-01	Wounds	10.5	November	New
1855-12-01	Wounds	5.0	December	New
1856-01-01	Wounds	0.5	January	New
1856-02-01	Wounds	0.0	February	New
1856-03-01	Wounds	0.0	March	New
1854-04-01	Other	7.0	April	Old
1854-05-01	Other	4.6	May	Old
1854-06-01	Other	2.5	June	Old
1854-07-01	Other	9.6	July	Old
1854-08-01	Other	11.9	August	Old
1854-09-01	Other	27.7	September	Old
1854-10-01	Other	50.1	October	Old
1854-11-01	Other	42.8	November	Old
1854-12-01	Other	48.0	December	Old
1855-01-01	Other	120.0	January	Old
1855-02-01	Other	140.1	February	Old
1855-03-01	Other	68.6	March	Old
1855-04-01	Other	21.2	April	New
1855-05-01	Other	12.5	May	New
1855-06-01	Other	9.6	June	New
1855-07-01	Other	9.3	July	New
1855-08-01	Other	6.7	August	New
1855-09-01	Other	5.0	September	New
1855-10-01	Other	4.6	October	New
1855-11-01	Other	10.1	November	New
1855-12-01	Other	7.8	December	New
1856-01-01	Other	13.0	January	New
1856-02-01	Other	5.2	February	New
1856-03-01	Other	9.1	March	New

```
DeathratePlot <- ggplot(data = rates,
             aes(x = factor(Date), y = Deaths, fill = Cause)) +
  geom_bar(width = 0.8, stat = "identity", position = "stack", color = "black") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        axis.ticks.x = element_blank(),
        legend.position = c(0.8, 0.8)) +
  scale_fill_brewer(type = "qual", palette = "Set1")

DeathratePlot
```



This bar chart makes it much easier to compare the height of each bar directly, something that's harder to do with the rose chart's petal shape. What jumps out immediately is the bell curve pattern in overall deaths, with the highest rates in the middle months. When we look specifically at the disease component (the largest portion of each bar), we can see a clear "before and after" effect when Nightingale's new practices were implemented. The visualization clearly shows that disease was responsible for most deaths and also saw the biggest reduction after the changes were made. This format complements Nightingale's original visualization by making the timeline clearer and allowing anyone, regardless of their data background, to quickly grasp both the pattern of deaths over time and the impact of the sanitation improvements. This makes it perfect for explaining the story to stakeholders who need to understand the bottom line without getting lost in complex visual formats.