

# Practice Tidying Data for ggplot2

Brandon Booth

Connor Crane

Amy Ly

Abraham Mendoza

Miles Moran

5/1/2022

## Problem 1: Tufte's Slopegraphs

```
library(tidyverse)
library(ggplot2)
```

```
#####
### Data Wrangling #####
#####

tax <- tribble(
  ~ Country,      ~ `1970`, ~ `1979`,
  "Sweden",       46.9,     57.4,
  "Netherlands",  44.0,     55.8,
  "Norway",       43.5,     52.2,
  "Britain",      40.7,     39.0,
  "France",       39.0,     43.4,
  "Germany",      37.5,     42.9,
  "Belgium",      35.2,     43.2,
  "Canada",       34.9,     35.8,
  "Finland",      34.9,     38.2,
  "Italy",        30.4,     35.7,
  "United States", 30.3,     32.5,
  "Greece",       26.8,     30.6,
  "Switzerland",  26.5,     33.2,
  "Spain",        22.5,     27.1,
  "Japan",        20.7,     26.6
)

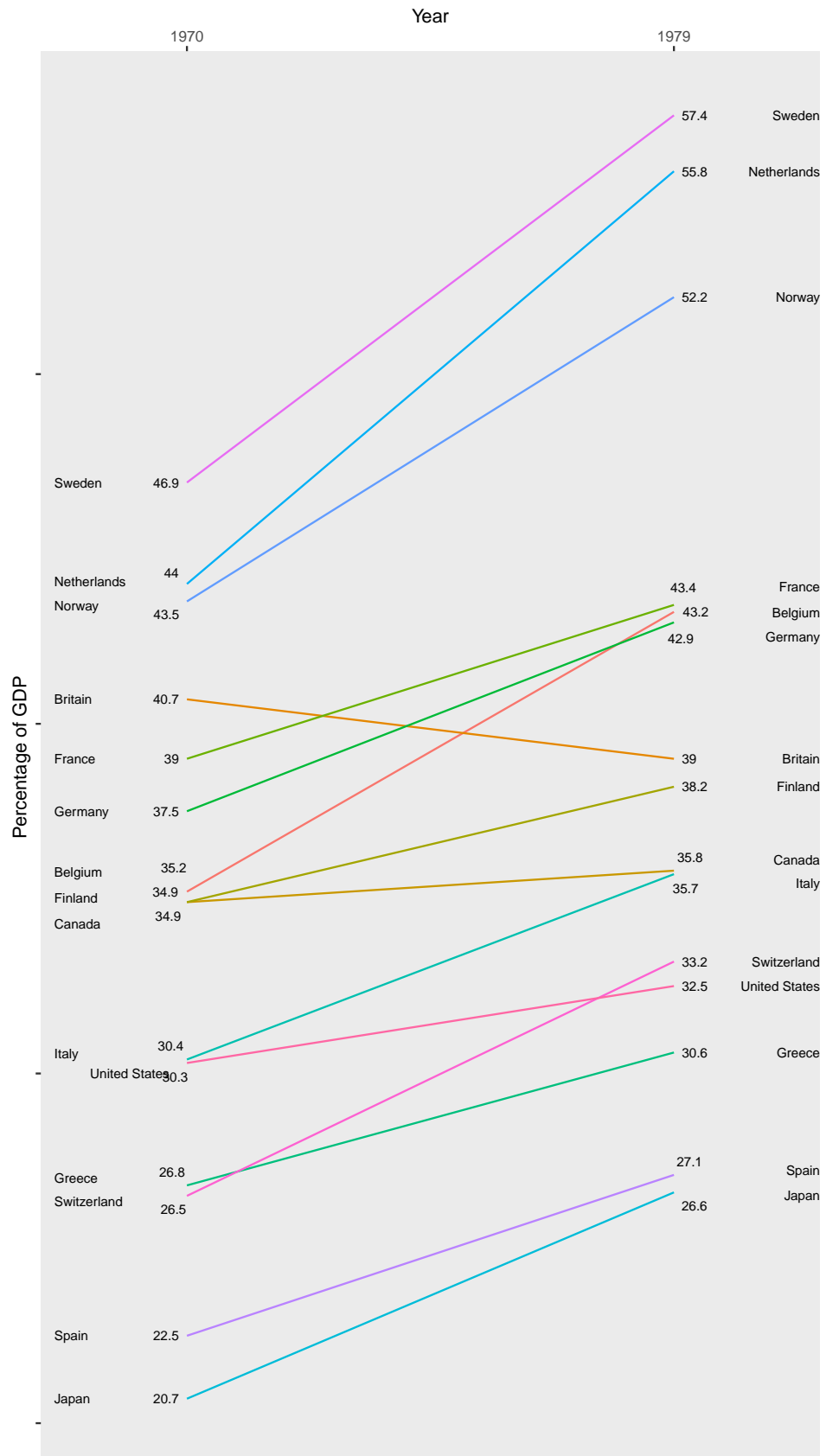
tax_new <- tax %>% pivot_longer(cols = c(`1970`, `1979`),
                                names_to = "year",
                                values_to = "taxpct")
```

```
#####
### Plotting #####
#####

ggplot(tax_new, aes(x = year, y = taxpct, label = Country)) +
  geom_line(aes(group = Country, color = Country)) +
  labs(main = "Receipts of Government as a Percentage of GDP",
        subtitle = "Comparison between 1970 and 1979",
        y = "Percentage of GDP",
        x = "Year") +
  scale_x_discrete(expand = expansion(mult=0.3),
                    position = "top") +
  theme(plot.margin = margin(0.1,0.1,0.05,0.05, "cm"),
        text = element_text(size = 10),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.text.y = element_blank(),
        legend.position = "none"
  ) +
  geom_text_repel(data = subset(tax_new, year == 1970),    # LHS Country Names
                  aes(label = Country),
                  nudge_x = -0.5,
                  size = 2.5,
                  hjust = 1,
                  segment.color = 'transparent'
  ) +
  geom_text_repel(data = subset(tax_new, year == 1970),    # LHS GDP% Numbers
                  aes(label = taxpct),
                  size = 2.5,
                  nudge_x = -0.02,
                  segment.color = 'transparent'
  ) +
  geom_text_repel(data = subset(tax_new, year == 1979),    # RHS Country Names
                  aes(label = Country),
                  nudge_x = 0.7,
                  size = 2.5,
                  hjust = 1,
                  segment.color = 'transparent'
  ) +
  geom_text_repel(data = subset(tax_new, year == 1979),    # RHS GDP% Numbers
                  aes(label = taxpct),
                  size = 2.5,
                  nudge_x = 0.02,
                  segment.color = 'transparent'
  )

```

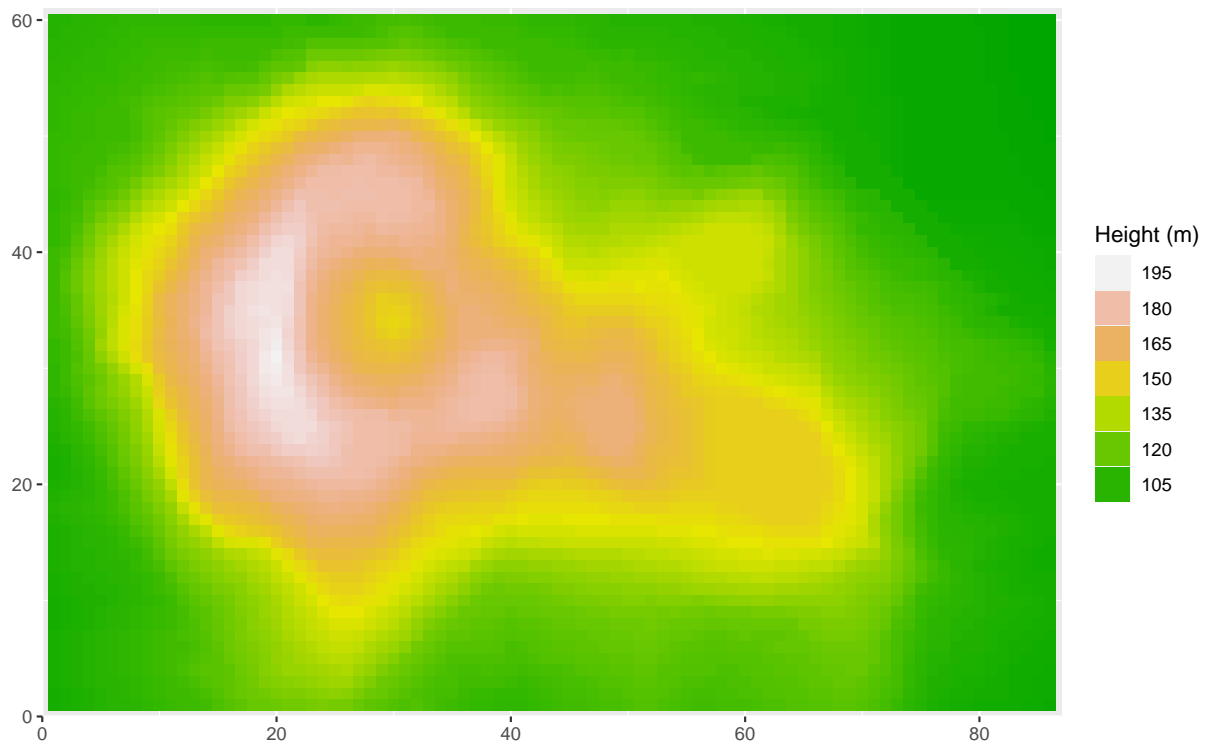
# Comparison between 1970 and 1979



## Problem 2: Heatmap of Maunga Whau (Mt Eden)

```
#####  
## Data Wrangling #####  
#####  
  
volcano_tbl <- as_tibble(volcano)  
  
colnames(volcano_tbl) <- 1:ncol(volcano)  
  
volcano_new <-  
  volcano_tbl %>%  
  mutate(y = 1:nrow(volcano)) %>%  
  pivot_longer(cols = 1:length(volcano_tbl),  
               names_to = "x",  
               values_to = "z") %>%  
  mutate(x = as.numeric(x))  
  
#####  
## Plotting #####  
#####  
  
cap <- "Note that the right-hand-side of the plot is the /north/ side of the volcano"  
  
ggplot(volcano_new, aes(y, x, fill=z)) +  
  geom_tile() +  
  scale_x_continuous(limits=c(0,87), expand=c(0,0)) +  
  scale_y_continuous(limits=c(0,61), expand=c(0,0)) +  
  scale_fill_gradientn(  
    colours = terrain.colors(102),  
    breaks = seq(195, 90, by=-15),  
    name = "Height (m)",  
    guide = "legend") +  
  labs(x = "", y = "",  
       title = "Topographic Heatmap of Maunga Whau (Mt Eden)",  
       subtitle = "Discretized into Tracts of 10m x 10m",  
       caption = cap  
  )
```

Topographic Heatmap of Maunga Whau (Mt Eden)  
Discretized into Tracts of 10m x 10m



Note that the right-hand-side of the plot is the /north/ side of the volcano

## Problem 3: Few's Deviation Analysis

```
library(scales)
library(directlabels)
library(grid)

budget <- tribble(
  ~ Expenses, ~ Jan, ~ Feb, ~ Mar, ~ Apr, ~ May, ~ Jun, ~ Jul, ~ Aug, ~ Sep, ~ Oct, ~ Nov, ~ Dec,
  "Domestic Actual", 84853, 84838, 88103, 85072, 88723, 90384, 89374, 95273, 94239, 92394, 96934, 105034,
  "Domestic Budget", 83000, 83830, 84668, 85515, 86370, 87234, 88106, 88987, 89877, 90776, 91684, 92600,
  "International Actual", 12538, 12438, 14934, 14033, 13945, 15938, 14086, 15934, 13945, 17338, 19384, 22394,
  "International Budget", 12000, 12600, 13860, 13200, 13860, 15246, 14520, 15246, 16771, 15972, 16771, 18448
)

#####
### Data Wrangling #####
#####

month_names <- colnames(budget)[-1]

budget_data <-
  budget %>%
  pivot_longer(cols = Jan:Dec,
               names_to = "Month",
               values_to = "Amount") %>%
  separate(Expenses,
           into = c('exp_loc', 'exp_type'),
           sep = ' ') %>%
  pivot_wider(names_from = exp_type,
              values_from = Amount) %>%
  mutate(exp_variance = (Actual - Budget)) %>%
  mutate(exp_variance_percent = (Actual - Budget)/Budget) %>%
  mutate(label = if_else(Month == 'Dec',
                        as.character(exp_loc),
                        NA_character_))

# Code used to remove clipping after creating a ggplot
my_plot_func <- function(p) {
  gt2 <- ggplotGrob(p)
  gt2$layout$clip[gt2$layout$name == "panel"] <- "off"
  grid.draw(gt2)
}
```

```
#####
## Plotting Figure 9.8 #####
#####
```

```
p1 <-
  ggplot(budget_data, aes(x = factor(Month), y = exp_variance,
                           group = exp_loc , color = exp_loc)) +
  geom_point() +
  geom_line() +
  theme_classic(10) +
  scale_x_discrete(limits = month_names) +
  scale_y_continuous(labels = function(x) format(x, big.mark = ","),
                    limits = c(-4000, 14000),
                    breaks = c(seq(-2000, 14000, 2000))) +
  scale_color_manual(values = c("#000000", "#808080")) +
  labs(x = "", y = "",
       title = "Expense Variance from Budget in U.S. Dollars",
       caption = "Figure 9.8"
  ) +
  theme(plot.title = element_text(face = "bold", hjust=0.5),
        legend.position = "none"
  ) +
  geom_dl(aes(label = exp_loc),
         method = list(dl.combine("last.points"))
  ) +
  theme(plot.margin = unit(c(1,10,1,1), "lines"))

my_plot_func(p1)
```

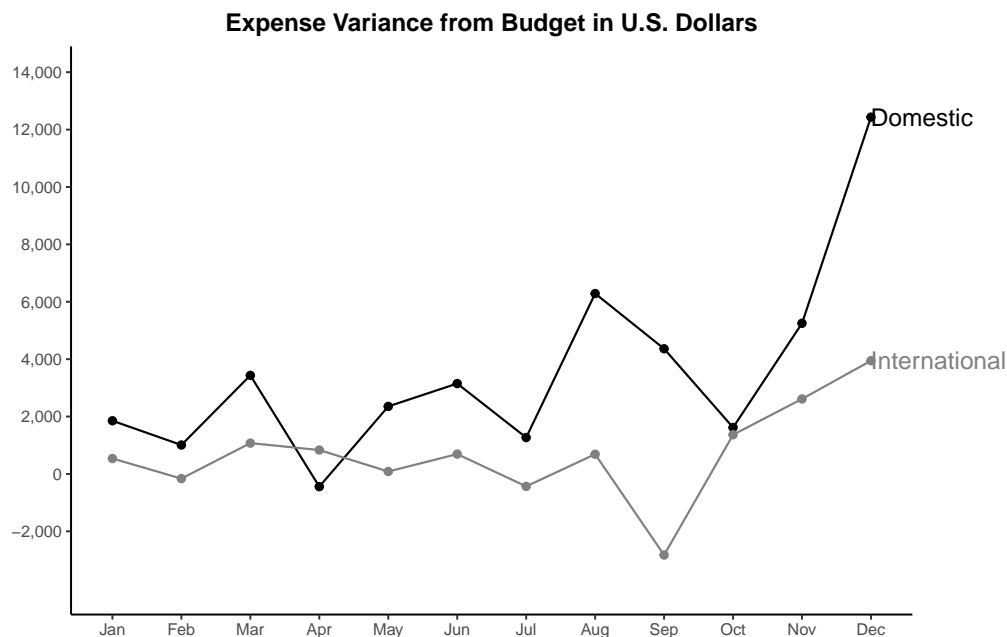


Figure 9.8

```
#####
### Plotting Figure 9.9 #####
#####
```

```
p2 <-
  ggplot(budget_data, aes(x = factor(Month), y = exp_variance_percent,
                           group = exp_loc , color = exp_loc)) +

  geom_point() +
  geom_line() +
  theme_classic(10) +
  scale_x_discrete(limits = month_names) +
  scale_y_continuous(labels = function(x) percent(x, accuracy = 1),
                     limits = c(-.2, .25),
                     breaks = c(seq(-.2, .25, .05))) +
  scale_color_manual(values = c("#000000", "#808080")) +
  labs(x = "", y = "",
       title = "Percentage Variance of Expenses from Budget",
       caption = "Figure 9.9"
  ) +
  theme(plot.title = element_text(face = "bold", hjust=0.5),
        legend.position="none"
  ) +
  geom_dl(aes(label = exp_loc),
         method = list(dl.combine("last.points"))
  ) +
  theme(plot.margin = unit(c(1,10,1,1), "lines"))

my_plot_func(p2)
```

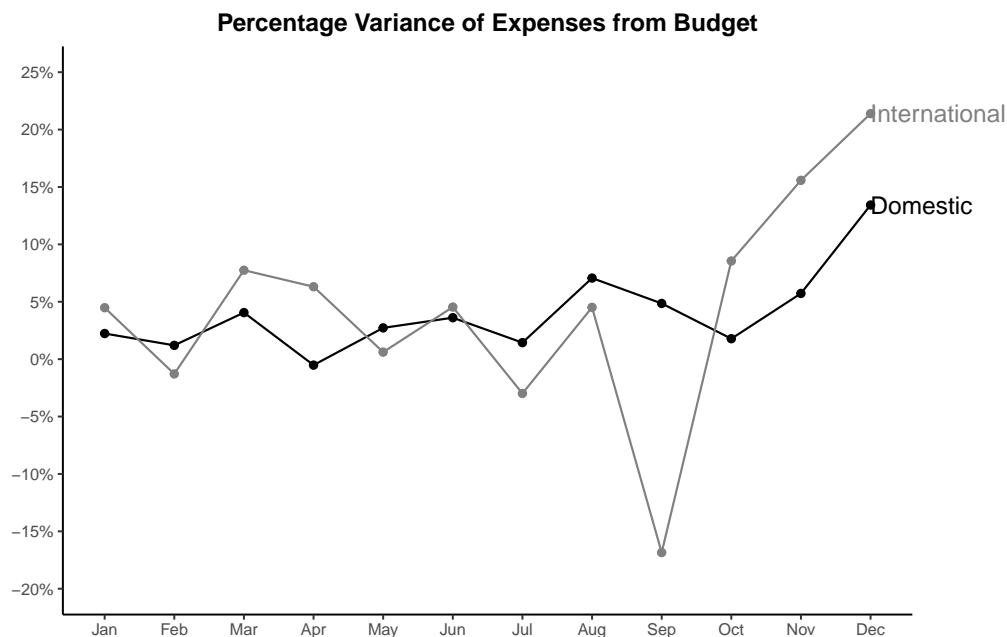


Figure 9.9