

Tidying data for ggplot2

Group 4

2025-05-03

Tufte slopegraph

```
# 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_tidy <- tax %>%
  pivot_longer(cols = c(`1970`, `1979`),
    names_to = "Year",
    values_to = "Tax Percent")

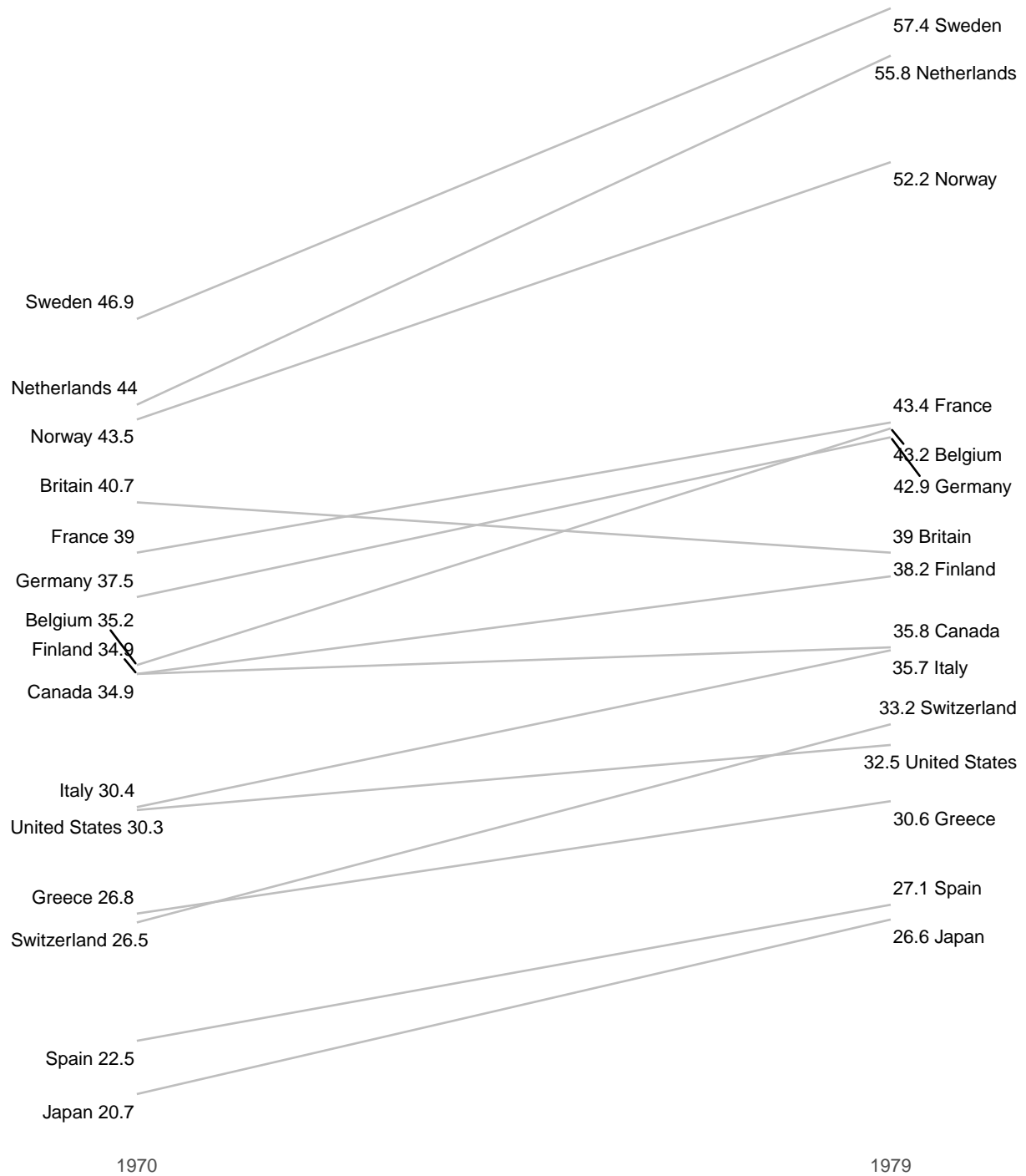
tax_tidy$Year <- as.numeric(tax_tidy$Year)

tax_tidy$`Tax Percent` <- round(tax_tidy$`Tax Percent`, 1)
```

Graphing

```
ggplot(tax_tidy, aes(x=Year, y=`Tax Percent`, group=Country)) +  
  geom_line(color = "grey") +  
  geom_text_repel(data=subset(tax_tidy, Year==1970),  
    aes(label=paste(Country, `Tax Percent`)),  
    direction = "y",  
    hjust = "right",  
    size=3) +  
  geom_text_repel(data=subset(tax_tidy, Year==1979),  
    aes(label=paste(`Tax Percent`, Country)),  
    direction = "y",  
    hjust = "left",  
    size=3) +  
  scale_x_continuous(breaks = c(1970, 1979), limits = c(1969, 1980)) +  
  theme_minimal() +  
  theme(axis.title = element_blank(),  
    axis.text.y = element_blank(),  
    axis.ticks = element_blank(),  
    panel.grid = element_blank()) +  
  labs(title = "Current Receipts of Government as a \nPercentage of Gross Domestic \nProduct, 1970 and 1979")
```

Current Receipts of Government as a Percentage of Gross Domestic Product, 1970 and 1979



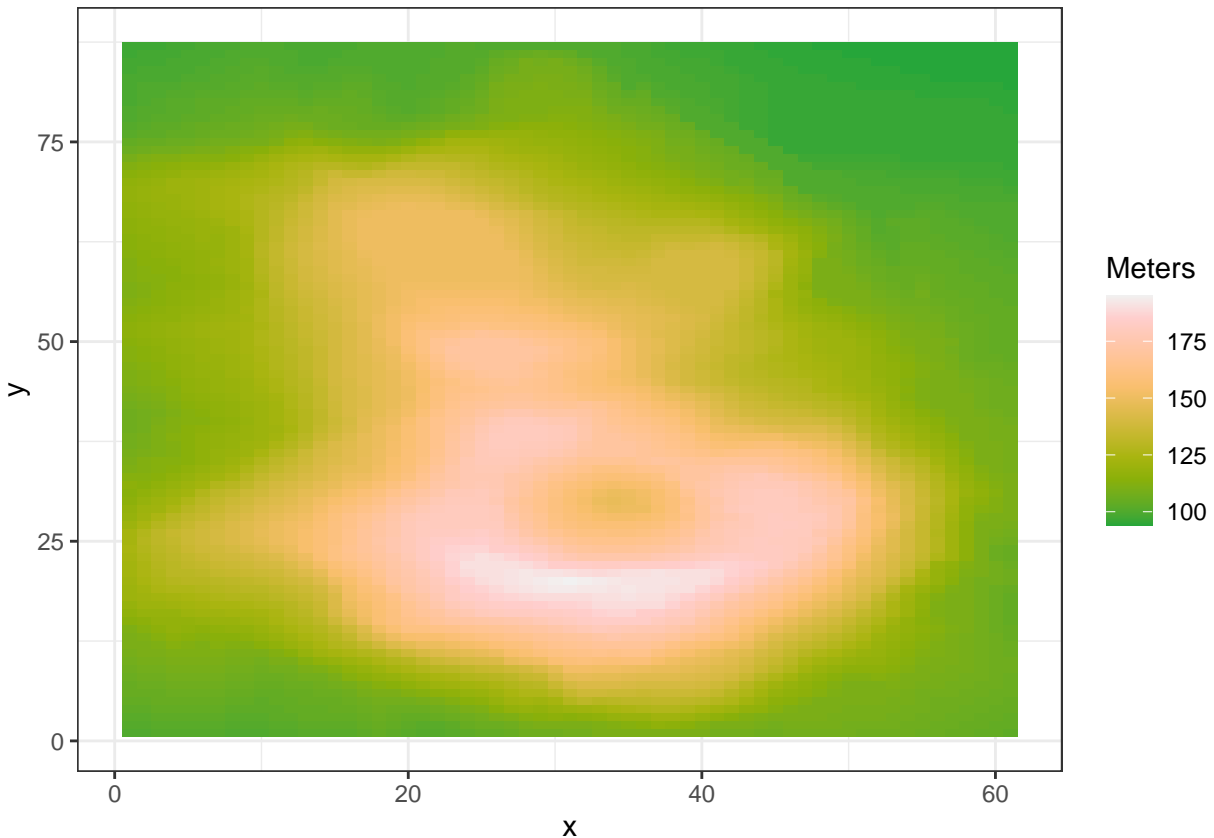
Heatmap of Maunga Whau

```
# Data wrangling
volcano_tbl <- as_tibble(volcano)

colnames(volcano_tbl) <- 1:ncol(volcano)
volcano_tbl$row <- 1:nrow(volcano_tbl)

volcano_long <- volcano_tbl %>%
  pivot_longer(
    cols = -row,
    names_to = "x",
    values_to = "z"
  ) %>%
  mutate(
    x = as.integer(x),
    y = row
  ) %>%
  select(x, y, z)
# Graphing
library(colorspace)

ggplot(volcano_long, aes(x = x, y = y, z = z)) +
  geom_raster(aes(fill = z)) +
  scale_fill_continuous_sequential("Terrain", rev = F, name = "Meters") +
  theme_bw()
```



Few's Deviation Analysis

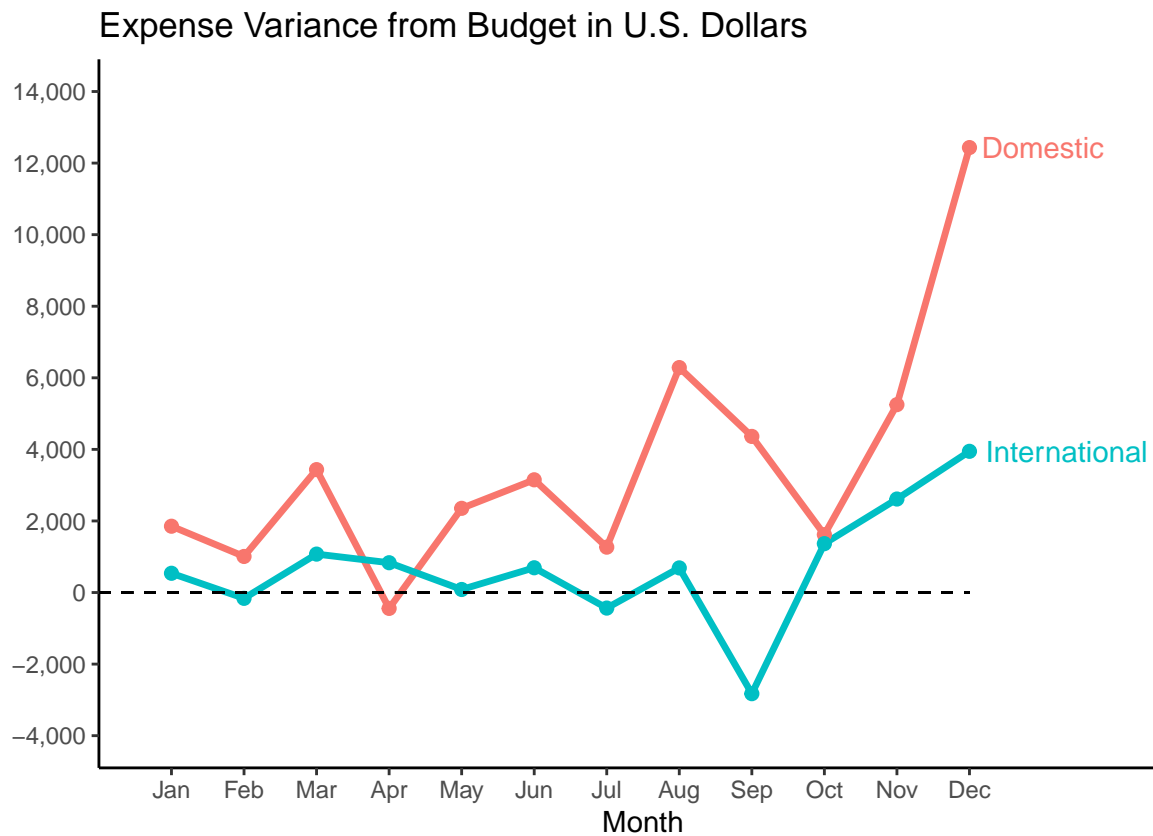
```
# Data wrangling
budget <- tribble(
  ~ Expenses,           ~ Jan, ~ Feb, ~ Mar, ~ Apr, ~ May, ~ Jun, ~ Jul, ~ Aug, ~ Sep, ~ Oct, ~ Nov,
  "Domestic Actual",    84853, 84838, 88103, 85072, 88723, 90384, 89374, 95273, 94239, 92394, 96934,
  "Domestic Budget",    83000, 83830, 84668, 85515, 86370,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

months <- paste(names(budget)[2:13])           # make vector of months

df.budget.graph <- budget |>
  pivot_longer(cols = months,                  # widen data to consolidate months
               names_to = "Month",
               values_to = "Budget") |>
  separate(Expenses,                          # separate Expenses based on space
           into = c("Nationality", "Spending"),
           sep = " ") |>
  pivot_wider(names_from = Spending,           # widen to position budget amounts
              values_from = Budget) |>
  mutate(Variance = Actual - Budget,          # calculate needed values
         Percent_deviation = (Variance/Budget), # modified this so we could display the % sign easily
         Month = factor(Month,                # factor month for easy graphing
                        levels = months))

# Graphing
# Replication of Figure 9.8
ggplot(df.budget.graph, aes(x = Month, y = Variance, color = Nationality, group = Nationality)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  scale_y_continuous(breaks = seq(-4000, 14000, by = 2000),
                    limits = c(-4000, 14000), # added the y range we wanted with addition of limits
                    labels = comma # from the scales package) +
  geom_text(
    data = subset(df.budget.graph, Month == "Dec"),
    aes(label = Nationality),
    hjust = -0.1,
    show.legend = FALSE
  ) +
  geom_segment(aes(x = 0, xend = 12, y = 0, yend = 0),
              inherit.aes = FALSE,
              linetype = "dashed",
              size = 0.5
  ) +
  labs(
    title = "Expense Variance from Budget in U.S. Dollars",
    x = "Month",
    y = "",
    color = "Nationality"
  ) +
  theme_classic() +
```

```
theme(legend.position = "none",
      plot.margin = unit(c(0.25, 0.25, 0.25, 0.25), "cm")) +
scale_x_discrete(expand = expansion(add = c(0.1, 3)))
```



```
# Replication of Figure 9.9
ggplot(df.budget.graph, aes(x = Month, y = Percent_deviation, color = Nationality, group = Nationality))
  geom_line(size = 1.2) +
  geom_point(size = 2) +

  scale_y_continuous(
    breaks = seq(-0.20, 0.25, by = 0.05), # numeric breaks
    limits = c(-0.20, 0.25), # added the y range we wanted with addition of limits
    labels = scales::percent # numbers to percentages
  ) +
  geom_text(
    data = subset(df.budget.graph, Month == "Dec"),
    aes(label = Nationality),
    hjust = -0.1,
    show.legend = FALSE
  ) +
  geom_segment(aes(x = 0, xend = 12, y = 0, yend = 0),
    inherit.aes = FALSE,
    linetype = "dashed",
    size = 0.5
  ) +
  labs(
```

```

title = "Percentage Variance of Expenses from Budget Budget",
x = "Month",
y = "",
color = "Nationality"
) +
theme_classic() +
theme(legend.position = "none",
      plot.margin = unit(c(0.25, 0.25, 0.25, 0.25), "cm")) +
scale_x_discrete(expand = expansion(add = c(0.1, 3)))

```

