Intro to ggplot2 - Ian Lyons

Ian Lyons

1/20/2019

Read in tax expenditure data for Plot 1

Source for tax expenditures on the Treasury Department Website: https://home.treasury.gov/policy-issues/tax-policy/tax-expenditures

```
base_df <- read_csv(file = 'tax_expenditures_18to28_renamed.csv')

budg_2019 <- read_csv(file = 'outlays-fy2019.csv')
```

Reshape Data For Plot 1

Create a long-form version of the tax expenditures dataset to use in a time-series plot. That is, create one row for each category, detail, and fiscal year. Then drop the cumulative 'FY2019_to_2028' column since I won't be using it.

```
tax_long <- gather(base_df, fiscal_year, expenditure, '2018':'2028', factor_key=TRUE)
tax_long$FY2019_to_2028 <- NULL
```

Plot 1

```
## Create a time-series dataset so that tax expenditures can be compared over the fiscal years included

time_series <- tax_long %>% select(fiscal_year, Category, expenditure) %>%
     group_by(fiscal_year, Category) %>%
     tally(expenditure)

## Convert the expenditure column to be in billions of dollars
time_series <- mutate(.data = time_series, amount_billions = n/1000)
time_series$n <- NULL
bar_to_plot <- ggplot(data=time_series, aes(x=fiscal_year, y=amount_billions, color=Category)) + geom_c

plot1 <- bar_to_plot +
    scale_y_continuous(name = "Tax Expenditures in Billions of Nominal Dollars", labels =c('$0', '$433'
                       , breaks=c(0, 433, 866, 1300, 1733, 2166, 2600), limits = c(0,2600)) +

    scale_x_discrete(name='Fiscal Year') +

    labs(title = "US Federal Government Tax Expenditures, 2018-2028",
         subtitle = "The United States spends trillions of dollars on tax loopholes each year",
         caption = "Source: United States. Department of Treasury. Tax Policy: Tax Expenditures \n
         * These estimates are made relative to current law as of July 1, 2018",
         x = "Fiscal Year" +

             theme(plot.title = element_text(color="black", size=14, face="bold", hjust = 0.5),
                   plot.subtitle = element_text(color="black", size=12, hjust =0.5),
                   axis.title.x = element_text(color="black", size=10),
                   axis.title.y = element_text(color="black", size=10),
                   plot.caption = element_text(color="black", size=8, face="italic")
```

```
                    )
    )
ggsave('tax_expenditures_to_2028.pdf', path="visuals/", width=18, height=9, units='in')
```

Read in Federal Budget Outlays data

Source for federal outlays on the Office of Management and Budget website: * https://www.whitehouse.gov/omb/supplemental-materials/

```
budg_2019 <- read_csv(file = 'outlays-fy2019.csv')
```

Reshape Data For Plot 2

Create a long-form version of the outlays dataset to use in a time-series plot. That is, create one row for each line item (agency name, bureau name, account name, etc.) and fiscal year.

```
budget_longform <- gather(budg_2019, fiscal_year, expenditure, '1962':'2018', factor_key=TRUE)

## Keep only descriptive column names rather than codes.
keep_cols <- c('Agency Name', 'Bureau Name', 'Account Name', 'Treasury Agency Code', 'Subfunction Title

budget_longform <- budget_longform[keep_cols]
```

Group by agency name and sum expenditures within the same agency and year.

```
outlays <- budget_longform %>% select(fiscal_year, `Agency Name`, expenditure) %>%
            group_by(fiscal_year, `Agency Name`) %>%
            tally(expenditure)

## Give the expenditures column a descriptive name
outlays <- mutate(.data = outlays, dollars_thousands = n)
outlays$n <- NULL

## Rename `Agency Name` in snake case.
colnames(outlays)[2] <- 'agency_name'
```