

# Data Exploration: U.S. City Daily High Temperatures

## Setting up code chunk specifications

- `echo = TRUE`: Code chunks will be included in the document unless specified in the specific code chunk (`echo = FALSE`)
- `warning = FALSE`: Warnings generated by code will not be included in the knitted document
- `message = FALSE`: Messages generated by code will not be included in the knitted document
- `root.dir = ""`: Sets the root directory for the document. When loading in data or images, the path will automatically be set to the root directory.

## Example code chunk

```
# Load in libraries
library(tidyverse)
library(stringr)

# Get specific date's historical TMINS and TMAXs
GetDayTemp = function(data) {
  current_day = format(Sys.Date(), "%m-%d")
  filtered_data = data %>%
    filter(grepl(current_day, date))
  return(filtered_data)
}

# Create a histogram for an input location
CreateTempHistogram = function(data, location) {
  current_day = Sys.Date()
  current_day = str_remove(current_day, "2023-")
  hist = ggplot(data) +
    geom_histogram(aes(TMAX), bins = 15) +
    theme_classic() +
    theme(plot.title = element_text(hjust = 0.5)) +
    ggtitle(location)

  return(hist)
}

# Read in temperature data for St. Paul and Urbana
stpaul_data = read.csv("Data/stpaul-temps.csv")

urbana_data = read.csv("Data/urbana-temps.csv")

# Combine St. Paul and Urbana data in a list to use lapply functions
data_list = list(stpaul_data, urbana_data)
```

```

# Change the date column to Type Date for summary statistics
data_list = lapply(data_list, mutate, date = as.Date(date))

# Create a variable that holds the current date in "Month Day" format to add to document text
today = format(Sys.Date(), "%b %d")

# Filter temperature data to only include days that mach the current day and month
clean_data = lapply(data_list, GetDayTemp)

# Create a histogram of maximum temperatures in the given location on today s date
hist_MN = CreateTempHistogram(clean_data[[1]], location = "St. Paul, MN")
hist_IL = CreateTempHistogram(clean_data[[2]], location = "Urbana, IL")

# Set the file name to save the histogram as
stpaul_filename = "stpaul-TMAX-hist.pdf"
urbana_filename = "urbana-TMAX-hist.pdf"

# Set the folder path for where to save the histogram
path = "Plots/"

```

## Saving plots using ggsave()

- ggsave() allows us to save plots with a specified filename and path.

```

# Save the histogram as a pdf
ggsave(filename = stpaul_filename,
       hist_MN,
       path = path)

ggsave(filename = urbana_filename,
       hist_IL,
       path = path)

```

## Summary Statistics

### Combining text and code

- We can combine text and values stored in variables using ‘ r ‘. This method is used in the paragraph below.

Maximum daily temperature records were taken in St. Paul, MN from 1938 - 2021. Maximum daily temperature records were taken in Urbana, IL from 1903 - 2021. The table below shows the summary statistics for historical maximum temperatures in Minnesota and Urbana, IL on May 30.

### Creating a summary table using kable

```

knitr::kable(data.frame(format(summary(clean_data[[1]]$TMAX)),
                        format(summary(clean_data[[2]]$TMAX))),
             col.names = c("St. Paul", "Urbana"))

```

	St. Paul	Urbana
Min.	57.02	57.02
1st Qu.	69.98	71.96
Median	75.92	78.98
Mean	75.44	78.53
3rd Qu.	80.96	84.92
Max.	95.00	93.02

## Combining multiple plots in a grid

- To access outside images, we can use `file.path()` to hold the path to the given image file and then render the image to our document using `knitr::include_graphics()`. In our code chunk header, we included `'figures-side'`, `'fig.show = "hold"'`, and `'out.width = "50%"'` in order to have our images rendered side-by-side.

```
hist_MN = file.path("Plots/stpaul-TMAX-hist.pdf")
hist_IL = file.path("Plots/urbana-TMAX-hist.pdf")

knitr::include_graphics(c(hist_MN, hist_IL))
```

