Assingment_datavis

Pavan sai

2025-10-07

Data Visualisation Homework - 1

```
sun <- read.csv("~/Downloads/sunshine.csv", stringsAsFactors = FALSE)
head(sun)
```

```
##    sunshine    city     lat       lon month monthnum
## 1        69 Seattle 47.6062 -122.332   Jan        0
## 2       108 Seattle 47.6062 -122.332   Feb        1
## 3       178 Seattle 47.6062 -122.332   Mar        2
## 4       207 Seattle 47.6062 -122.332   Apr        3
## 5       253 Seattle 47.6062 -122.332   May        4
## 6       268 Seattle 47.6062 -122.332   Jun        5
```

```
unique(sun$city)
```

```
## [1] "Seattle"        "Phoenix"        "New York"       "Chicago"
## [5] "Houston"        "Miami"          "Salt Lake City"
```

```
names(sun)[names(sun) == "sunshine"] <- "avg_sunshine_hours"

# Convert month to full name (robust way: use monthnum 0..11)
sun$month <- factor(month.name[sun$monthnum + 1], levels = month.name)
```

```
temp <- data.frame(
  city = rep(c("Seattle", "Phoenix", "New York", "Chicago",
               "Houston", "Miami", "Salt Lake City"), each = 12),
  month = rep(month.name, 7),
  avg_temp = c(
    # Seattle
    42, 45, 48, 52, 58, 63, 68, 69, 64, 55, 48, 43,
    # Phoenix
    56, 60, 65, 73, 83, 92, 96, 95, 89, 77, 64, 56,
    # New York
    33, 36, 44, 55, 65, 75, 80, 78, 72, 61, 50, 39,
    # Chicago
    28, 32, 41, 53, 63, 73, 78, 76, 69, 57, 43, 32,
    # Houston
    53, 57, 63, 69, 76, 82, 84, 84, 80, 72, 63, 55,
    # Miami
```

1

```
      68, 70, 73, 76, 80, 83, 85, 85, 84, 81, 76, 71,
      # Salt Lake City
      30, 36, 45, 52, 61, 72, 81, 79, 70, 56, 43, 32
    )
)
```

```
# Merge
merged <- merge(sun[, c("city", "month", "avg_sunshine_hours")],
                temp, by = c("city", "month"))
```
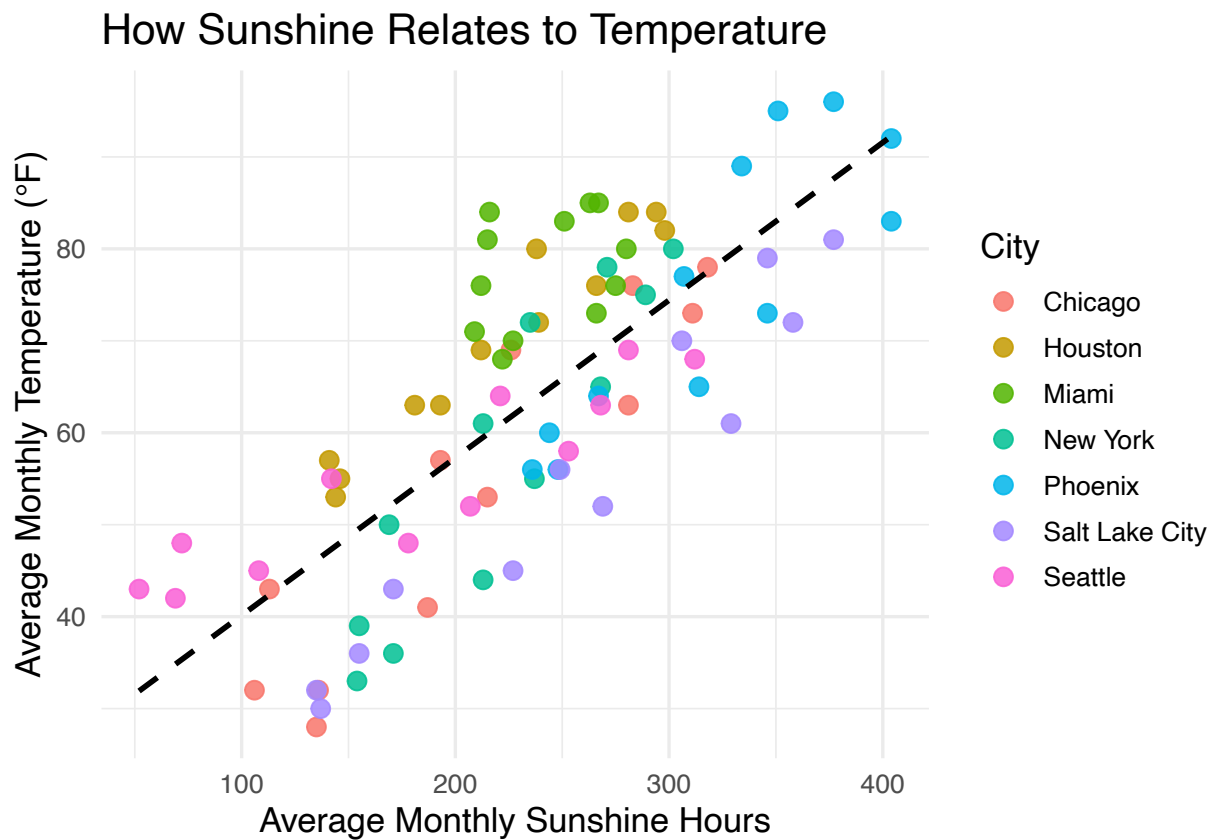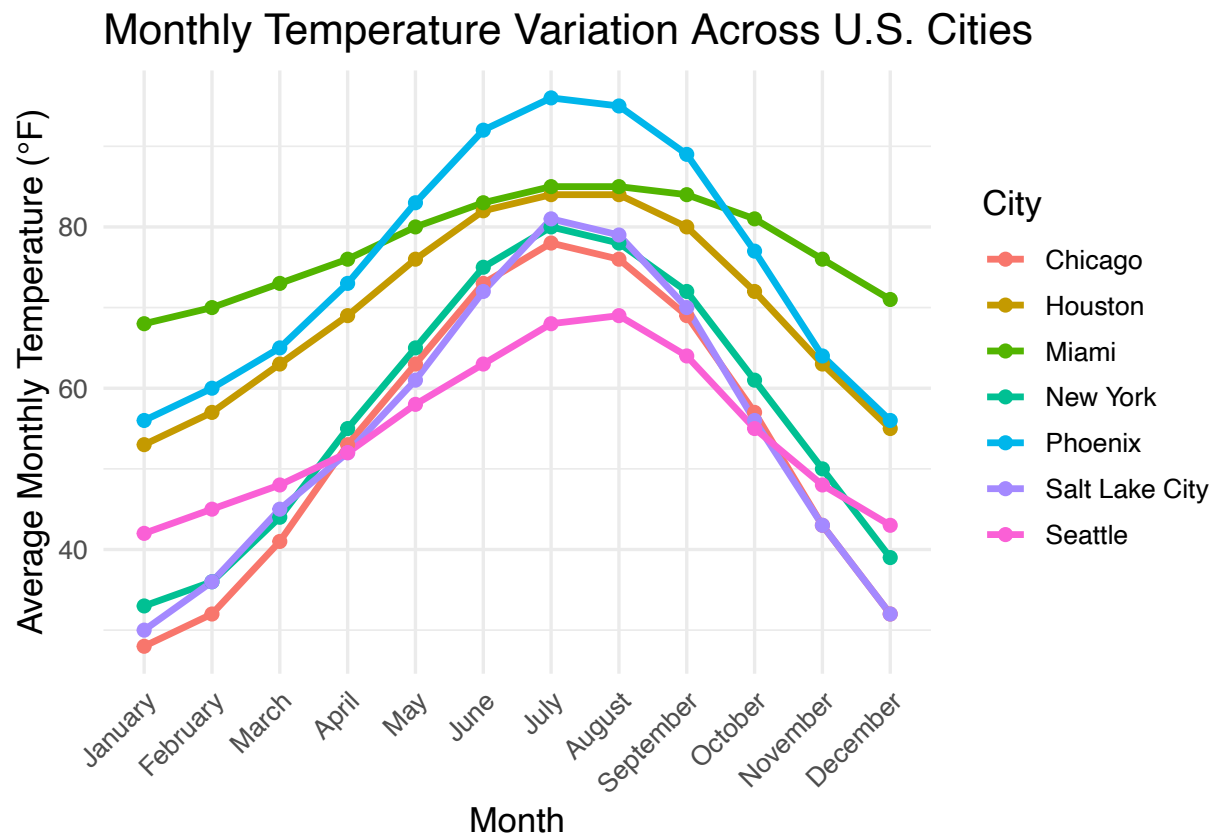
```
library(ggplot2)
ggplot(merged, aes(x = avg_sunshine_hours, y = avg_temp, color = city)) +
  geom_point(size = 3, alpha = 0.85) +
  geom_smooth(method = "lm", se = FALSE, color = "black", linetype = "dashed") +
  labs(
    title = "How Sunshine Relates to Temperature",
    x = "Average Monthly Sunshine Hours",
    y = "Average Monthly Temperature (°F)",
    color = "City"
  ) +
  theme_minimal(base_size = 13)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
ggplot(merged, aes(x = month, y = avg_temp, group = city, color = city)) +
  geom_line(linewidth = 1.2) +
  geom_point(size = 2) +
  labs(
    title = "Monthly Temperature Variation Across U.S. Cities",
    x = "Month",
    y = "Average Monthly Temperature (°F)",
    color = "City"
  ) +
  theme_minimal(base_size = 13) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


I wanted to explore how sunshine and temperature are related across different U.S. cities and how that relationship changes throughout the year. In the first plot, I compared the average monthly sunshine hours with the average monthly temperature. Each point represents one city in one month, and the color shows which city it belongs to. I added a dashed trend line using a simple linear model to see if there was a general pattern. The upward slope made it clear that places with more sunshine, like Phoenix and Miami, usually stay warmer, while cities with fewer sunny hours, like Seattle, tend to be cooler. This confirmed my initial assumption that sunshine and temperature rise together.

For the second plot, I wanted to see how temperature changes over time for each city, so I mapped the months on the x-axis and average temperature on the y-axis, using colored lines for each city. The line geometry makes the seasonal pattern easy to notice—northern cities like Chicago and New York have sharp peaks and dips, while southern cities like Houston and Miami stay fairly consistent. Together, these two visuals tell a simple but clear story: cities with more sunlight generally experience warmer weather, and geography plays a big role in how temperature changes across the seasons.