

Git Set Go

AUTHORS

Siddhi Jadhav

Sanch Anand

Thi Lan Anh Tran

PUBLISHED

May 27, 2025

1 Loading Libraries and Reading the Data

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

```
data <- read_csv("data/weatherAUS.csv")
```

2 Dropping high-missing columns, Converting 'Date' to Date type, Dropping rows where 'RainTomorrow' is missing

```
data <- data %>%
  select(-Evaporation, -Sunshine, -Cloud9am, -Cloud3pm)

data <- data %>%
  mutate(Date = ymd(Date))

data <- data %>%
  filter(!is.na(RainTomorrow))
```

3 Dropping wind direction columns, Converting RainToday and RainTomorrow to binary 0/1 and Dropping any remaining NAs

```
data <- data %>%
  select(-WindGustDir, -WindDir9am, -WindDir3pm)

data <- data %>%
  mutate(
    RainToday = if_else(RainToday == "Yes", 1, 0),
    RainTomorrow = if_else(RainTomorrow == "Yes", 1, 0)
  )

data <- data %>% drop_na()
```

4 Saving the cleaned Data

```
write_csv(data, "data/data_cleaned.csv")
```

5 Results

Figure [Figure 1](#) displays the 3pm temperature distribution by observation count for the top 10 Australian cities. The tropical, desert, and temperate regions of Australia are clearly distinguished by this research.

Tropical and desert climate conditions are reflected in the highest and most stable afternoon temperatures found in cities like **Darwin** and **Alice Springs**. Smaller interquartile ranges are seen in these cities, suggesting hot, consistent weather that is typical of the area. Conversely, cities with temperate, coastal climates, such as **Hobart**, **Melbourne**, and **Sydney**, show higher variability and cooler median temperatures. Cities like **Perth** and **Brisbane** have outliers, indicating sporadic excessive temperature events that may be caused by particular weather systems.

All things considered, the boxplot illustrates how Australia's varied topography affects temperature distributions. While southern and coastal towns exhibit milder and more diverse climates, northern cities are often hotter and more stable. These findings show how latitude, altitude, and proximity to the coast affect temperature trends and are consistent with the anticipated climatic variation throughout Australia.

```
top10 <- data %>%
  count(Location, sort = TRUE) %>%
  slice_head(n = 10) %>%
  pull(Location)

data %>%
  filter(Location %in% top10) %>%
  ggplot(aes(x = Location, y = Temp3pm)) +
  geom_boxplot(fill = "pink") +
  labs(x = "City", y = "Temperature at 3pm (°C)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

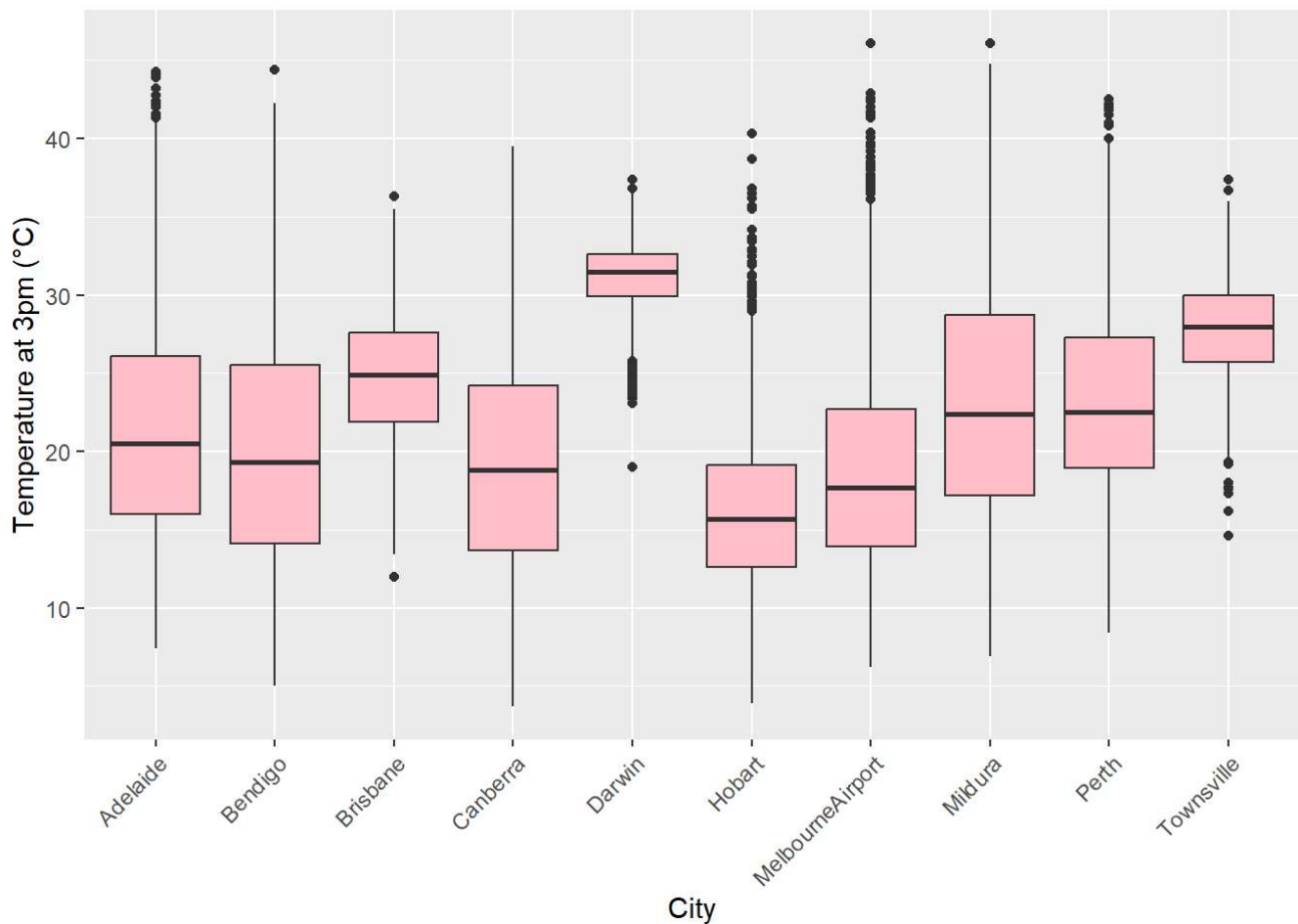


Figure 1: Distribution of 3pm temperatures across the top 10 Australian cities by observation count, highlighting differences between tropical and temperate regions.

As shown in [Figure 1](#) cities such as Darwin and Alice Springs experience the highest and most consistent afternoon temperatures. On the other hand, Melbourne and Hobart show cooler and more varied temperature distributions. This supports the expected climatic variation between Australia's tropical north and temperate south.