```
# Load necessary libraries
#tidyverse used for data manipulation, visualization, and analysis.
#readr is used for reading csv files
#readr: A package within tidyverse used for reading CSV files.

library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.5.1     ✓ tibble    3.2.1
## ✓ lubridate 1.9.4     ✓ tidyr     1.3.1
## ✓ purrr     1.0.4
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to bec
ome errors
```

```
library(readr)

# Step 1: Read CSV file, note there are no column names
#we also store the data in "arrival_data"
arrival_data <- read_csv("arrival_delays.csv", col_names = FALSE)
```

```
## Rows: 6 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (7): X1, X2, X3, X4, X5, X6, X7
## lgl (1): X8
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Step 2: Assign proper column names
#the datset includes the airline name, the status of the flights (on-time or delayed, the citi
es that the planes flew out of and an extra column that i need to get rid of later)

colnames(arrival_data) <- c("Airline", "Status", "Los_Angeles", "Phoenix", "San_Diego", "San_F
rancisco", "Seattle", "Extra_Column")

# Step 3: Remove the extra column ( I kept having an NA column which messed up my result) Keep
s only the first 7 columns, discarding the unnecessary Extra_Column.
arrival_data <- arrival_data[, 1:7]  # Keep only valid columns

# ✅ Step 4: Remove completely empty rows BEFORE pivoting
#Removes rows where all columns contain NA (missing values).
```