COGS 137: Weather App Usage

Arundhati Calambur, Josh Wang, Nathan Ng

Introduction

Weather plays a significant role in our daily lives, from influencing our daily activities to impacting agriculture, transportation, and even our safety. In recent years, the effects of climate change have become increasingly apparent, and extreme weather events have become more frequent and severe. In this context, staying informed about weather conditions has become more crucial than ever before.

This report investigates the weather checking habits of Americans, aiming to understand how often people check the weather, what methods they use to get weather updates, and any regional or generational differences in weather checking habits. By analyzing this data, we can gain insights into how Americans stay informed about weather conditions, and identify areas where we can improve weather communication and education.

The report also examines the regional and generational differences in weather checking habits, highlighting how factors such as location and age can influence how people stay informed about weather conditions.

Overall, this report provides valuable insights into the weather checking habits of Americans, shedding light on how we can improve weather communication and education to better prepare for extreme weather events and mitigate the impacts of climate change.

Using data collected from a SurveyMonkey survey, adults from across America were asked whether they typically checked a daily weather report, how they checked the weather, and some information about their demographics. We used the data collected to answer the following questions.

Questions

- Is there a relationship between whether a person typically checks their weather app and their sex?
- Does weather app use differ based on where the respondent is located?
- Is there a relationship between household income and weather app?
- What is the most popular website/app used to check the weather?
- Can we predict whether a person typically checks the daily weather report based on the method they use to check the weather and their demographic information?

Load Packages

```
library(tidyverse)
library(tidymodels)
library(skimr)
library(viridis)
library(kableExtra)
```

The Data

This dataset was taken from a fivethirtyeight article that analyzed how people around America checked the weather and how often they checked the weather. The data was gathered using SurveyMonkey Audience to run the survey and collect responses. It ran from April 6, 2015 to April 10, 2015 and had 938 respondents, asking two main questions: How do you check the weather and do you check a weather report every day?

Additionally, the survey also collected demographic information on the respondents, including their gender, age, location, and total household income.

Data Import

The dataset was downloaded from the fivethirtyeight github repo and we load the csv file in to wrangle the data.

```
weather <- read.csv(file = "data/raw/weather_check.csv")
```

Data Wrangling

Before we start wrangling the data, we first take a look at the data and the data types we have. Our data set includes 9 columns, including id, whether the respondent checks the daily weather report, how they check it, and some demographic information like gender and region. All of the columns are categorical, with the exception of id which we won't be using in our analysis.

```
# Check dataset
glimpse(weather)
```

```
## Rows: 928
## Columns: 9
## $ RespondentID
## $ Do.you.typically.check.a.daily.weather.report.
## $ How.do.you.typically.check.the.weather.
## $ A.specific.website.or.app..please.provide.the.answer.
## $ If.you.had.a.smartwatch..like.the.soon.to.be.released.Apple.Watch...how.likely.or.unlikely.would.y
## $ Age
## $ What.is.your.gender.
## $ How.much.total.combined.money.did.all.members.of.your.HOUSEHOLD.earn.last.year.
## $ US.Region
```

We first start with making the column names more readable and simplify the question columns to concise statements describing the value of the data. We replace the questions with whether they check the daily weather report, or check daily weather for short. We also shorten how respondents typically check the weather and if they have a specific site or app to method of checking weather and other website or app since most of the responses were other options that weren't present in method of checking the weather question. Additionally, we shorten the smartwatch app to check weather on smartwatch.

```
# Update column names to readable fashion
weather_names <- c(
  "id",
  "Checks daily weather",
  "Method of checking weather",
  "Other website or app",
  "Check weather on smartwatch",
  "Age",
  "Gender",
  "Household income",
  "Region"
)

names(weather) <- weather_names
```

Now that we've made the columns more concise and readable, we explore all the unique values of each column to see what data we have and identify and potential data errors. Most of the data came from a selection of predefined options, so many of the columns have a limited number of unique values. However, we do notice

that many of the columns include hyphen symbols which is used as a missing value filler. We also noticed that the column containing the other website or apps data has a significant number of unique values as a result of this question being the only free response answer.

```
# Explore each column, checking for their values
uniq_vals <- weather |>
  select(-id) |>
  lapply(unique)

uniq_vals
```

```
## $`Checks daily weather`
## [1] "Yes" "No"
##
## $`Method of checking weather`
## [1] "The default weather app on your phone"
## [2] "A specific website or app (please provide the answer)"
## [3] "The Weather Channel"
## [4] "-"
## [5] "Internet search"
## [6] "Local TV News"
## [7] "Newspaper"
## [8] "Radio weather"
## [9] "Newsletter"
##
## $`Other website or app`
##   [1] "-"
##   [2] "Iphone app"
##   [3] "AccuWeather App"
##   [4] "nice"
##   [5] "Weather.com"
##   [6] "Weatherbug"
##   [7] "weather channel app"
##   [8] "Yahoo weather iphone"
##   [9] "Weather Puppy"
##  [10] "AccuWeather"
##  [11] "Apple weater"
##  [12] "The Weather Channel app"
##  [13] "google \"weather\""
##  [14] "Weather Underground"
##  [15] "Weather Timeline (android)"
##  [16] "1 weather"
##  [17] "Weather Channel app"
##  [18] "Weather Underground, also local tv news half the time."
##  [19] "weatherbug app"
##  [20] "weather bug"
##  [21] "FancyClock app on my phone and accuweather.com"
##  [22] "GOES West Satalite"
##  [23] "Intellicast / Storm"
##  [24] "ipod weather "
##  [25] "weatherbug"
##  [26] "weatherbug.com"
##  [27] "iphone weather app"
##  [28] "Weather Bug"
##  [29] "Accuweather"
```

```
##  [30] "www.wunderground.com"
##  [31] "The Weather Channel"
##  [32] "noaa.gov"
##  [33] "1weather"
##  [34] "yahoo"
##  [35] "Weather Bug App"
##  [36] "Yo Window"
##  [37] "Weather bug"
##  [38] "Dark Sky"
##  [39] "aol"
##  [40] "NOAA and The Weather Channel"
##  [41] "Weather Kitty"
##  [42] "Internet, tv, radio, newspaper"
##  [43] "App on Iphone"
##  [44] "yr.no"
##  [45] "weather.com"
##  [46] "weatherunderground"
##  [47] "wunderground.com"
##  [48] "Weather channel app"
##  [49] "intellicast"
##  [50] "Weather Risk"
##  [51] "local newspaper online  Weather Bug  NOAH"
##  [52] "AccuWeather or Weather Underground"
##  [53] "Weatherbug, Storm"
##  [54] "look/feel the weather outside"
##  [55] "accuweather"
##  [56] "apple-provided site"
##  [57] "Weather channel app on my phone and ipad"
##  [58] "Numerous local weather apps"
##  [59] "The Weather Channel app on my phone"
##  [60] "Storm Team 4"
##  [61] "KCRA online"
##  [62] "yahoo weather"
##  [63] "weatherbug on my phone"
##  [64] "the weather channel app on my phone"
##  [65] "Weather Channel.com"
##  [66] "Weather Channel app on phone"
##  [67] " Weather underground "
##  [68] "IPhone"
##  [69] "Weather Channel app android"
##  [70] "NOAA app"
##  [71] "Via my iPhone builtin app"
##  [72] "weatherforyou.com"
##  [73] "Weather Underground through HD Widgets"
##  [74] "weather app"
##  [75] "iPhone app"
##  [76] "National Weather Service site"
##  [77] "Weather channel"
##  [78] "the weather channel"
##  [79] "iPhone weather app"
##  [80] "The Weather Channel for iPhone or iPad "
##  [81] "m.accuweather.com"
##  [82] "National Weather Service"
##  [83] "weather channel"
```

```
##  [84] "Google App"
##  [85] "The Weather Channel app on my phone."
##  [86] "WEatherbug"
##  [87] "DirecTv Weather App"
##  [88] "iphone"
##  [89] "smart phone"
##  [90] "Weather app"
##  [91] "basic weather app on my iPhone"
##  [92] "phone"
##  [93] "WeatherBug"
##  [94] "the weather channel app"
##  [95] "Bing"
##  [96] "Weather Channel App"
##  [97] "Talk to my mother."
##  [98] "i look outside"
##  [99] "Weather Channel App in IPhone"
## [100] "Weather Channel"
## [101] "The weather app on the iPhone "
## [102] "The weather channel iPad app"
## [103] "weather underground"
## [104] "My husband usually informs me about the next day's weather."
## [105] "Default iPhone app"
## [106] "Google now"
## [107] "Chrome app"
## [108] "Apple Weather App"
## [109] "Iphone weather app"
## [110] "desktop icon"
## [111] "The one that comes on your iphone"
## [112] "the weather channel and local TV kxan"
## [113] "Wunderground"
## [114] "Intellicast (site) & WUnderground (app) "
## [115] "myphone"
## [116] "weather.gov"
## [117] "my phone has a pre-loaded app"
## [118] "www.weather.gov"
## [119] "Weather & Clock widget"
## [120] "yahoo weather app"
## [121] "Www.weather.com"
## [122] "weather app on iphone"
##
## $`Check weather on smartwatch`
## [1] "Very likely"       "Somewhat likely"   "Very unlikely"
## [4] "-"                 "Somewhat unlikely"
##
## $Age
## [1] "30 - 44" "18 - 29" "-"       "45 - 59" "60+"
##
## $Gender
## [1] "Male"   "-"      "Female"
##
## $`Household income`
##  [1] "$50,000 to $74,999"   "Prefer not to answer" "$100,000 to $124,999"
##  [4] "$150,000 to $174,999" "$25,000 to $49,999"   "-"
##  [7] "$0 to $9,999"         "$10,000 to $24,999"   "$75,000 to $99,999"
```

```
## [10] "$200,000 and up"      "$175,000 to $199,999" "$125,000 to $149,999"
##
## $Region
##  [1] "South Atlantic"     "-"                    "Middle Atlantic"
##  [4] "West South Central" "Pacific"              "West North Central"
##  [7] "East North Central" "Mountain"             "New England"
## [10] "East South Central"
```

To begin our wrangling, we first start with checking each column and converting its data type for our analysis. The Checks daily weather column tells us whether the respondent typically checks the daily weather report and will be used later in our analysis and models. We convert the values from Yes and No to True and False and convert those values into factors so we can make predictions using our logistic regression.

```
# Turn Checks daily weather to logical factor
weather <- weather |>
  mutate(`Checks daily weather` = case_when(
    `Checks daily weather` == "Yes" ~ TRUE,
    .default = FALSE
  ),
  `Checks daily weather` = as.factor(`Checks daily weather`))
```

The next column we wrangle is the Method of checking weather column. We noticed that many of the options could be more concise, so we turned each value into a shorter, more concise version of the same value. We changed the option of a specific website or app to Other because most of the values in the next column corresponding to the provided answers were often either options in this column or some other option that was not available in this current column.

```
# Turn Method of checking weather into readable categories
methods <- c("Default weather app", "The Weather Channel", "Internet", "Newspaper", "Newsletter", "Othe

weather <- weather |>
  mutate(
    `Method of checking weather` = case_when(
      `Method of checking weather` == "The default weather app on your phone" ~ "Default weather app",
      `Method of checking weather` == "The Weather Channel" ~ "The Weather Channel",
      `Method of checking weather` == "Internet search" ~ "Internet",
      `Method of checking weather` == "Newspaper" ~ "Newspaper",
      `Method of checking weather` == "Newsletter" ~ "Newsletter",
      `Method of checking weather` == "A specific website or app (please provide the answer)" ~ "Other"
      `Method of checking weather` == "Local TV News" ~ "TV",
      `Method of checking weather` == "Radio weather" ~ "Radio",
      .default = `Method of checking weather`
    )
  )
```

Next, we wrangle the free response data in the Other website or app column. We have more than a hundred different unique values, with many duplicates of the same idea but differ in how the respondent phrased it, capitalized it, or spaced it. We first remove all trailing white space and convert each value to lower case. This lowers the number of unique values we now have. Then, we went through each unique value and either standardized it such that values of similar nature now have the same value. For example, "weatherunderground", "wunderground", and "wunderground.com" were all converted to Weather Underground. Additionally, we found that many of the responses could have been a different response in the Method of checking weather column. This was especially prevalent with the default app on phones and The Weather Channel, as respondents should have selected a different option in the Method of checking weather column and many of them phrased their responses in different ways resulting in a high number of unique values that meant the same thing. To resolve this, we set these similar responses to Default weather app and

The Weather Channel respectively, similar to how we encoded the values in the other column.

```
# Standardize other websites or apps column

# Check unique values in data set
weather |>
  select("Other website or app") |>
  distinct()
```

```
##                                          Other website or app
## 1                                                           -
## 2                                                  Iphone app
## 3                                              AccuWeather App
## 4                                                        nice
## 5                                                 Weather.com
## 6                                                  Weatherbug
## 7                                           weather channel app
## 8                                          Yahoo weather iphone
## 9                                                 Weather Puppy
## 10                                                  AccuWeather
## 11                                                Apple weater
## 12                                      The Weather Channel app
## 13                                             google "weather"
## 14                                          Weather Underground
## 15                                 Weather Timeline (android)
## 16                                                  1 weather
## 17                                          Weather Channel app
## 18       Weather Underground, also local tv news half the time.
## 19                                              weatherbug app
## 20                                                 weather bug
## 21               FancyClock app on my phone and accuweather.com
## 22                                           GOES West Satalite
## 23                                          Intellicast / Storm
## 24                                                ipod weather
## 25                                                   weatherbug
## 26                                              weatherbug.com
## 27                                           iphone weather app
## 28                                                  Weather Bug
## 29                                                  Accuweather
## 30                                         www.wunderground.com
## 31                                          The Weather Channel
## 32                                                    noaa.gov
## 33                                                    1weather
## 34                                                       yahoo
## 35                                              Weather Bug App
## 36                                                    Yo Window
## 37                                                 Weather bug
## 38                                                    Dark Sky
## 39                                                         aol
## 40                                 NOAA and The Weather Channel
## 41                                                Weather Kitty
## 42                               Internet, tv, radio, newspaper
## 43                                                App on Iphone
## 44                                                        yr.no
## 45                                                  weather.com
```

```
## 46                                    weatherunderground
## 47                                      wunderground.com
## 48                                   Weather channel app
## 49                                           intellicast
## 50                                          Weather Risk
## 51                  local newspaper online  Weather Bug  NOAH
## 52                         AccuWeather or Weather Underground
## 53                                   Weatherbug, Storm
## 54                        look/feel the weather outside
## 55                                           accuweather
## 56                                 apple-provided site
## 57                  Weather channel app on my phone and ipad
## 58                        Numerous local weather apps
## 59                      The Weather Channel app on my phone
## 60                                       Storm Team 4
## 61                                        KCRA online
## 62                                      yahoo weather
## 63                                weatherbug on my phone
## 64                      the weather channel app on my phone
## 65                                   Weather Channel.com
## 66                           Weather Channel app on phone
## 67                                   Weather underground
## 68                                               IPhone
## 69                           Weather Channel app android
## 70                                            NOAA app
## 71                             Via my iPhone builtin app
## 72                                   weatherforyou.com
## 73                    Weather Underground through HD Widgets
## 74                                       weather app
## 75                                        iPhone app
## 76                           National Weather Service site
## 77                                    Weather channel
## 78                                  the weather channel
## 79                                   iPhone weather app
## 80                    The Weather Channel for iPhone or iPad
## 81                                  m.accuweather.com
## 82                               National Weather Service
## 83                                    weather channel
## 84                                         Google App
## 85                       The Weather Channel app on my phone.
## 86                                       WEatherbug
## 87                                  DirecTv Weather App
## 88                                               iphone
## 89                                       smart phone
## 90                                       Weather app
## 91                            basic weather app on my iPhone
## 92                                               phone
## 93                                        WeatherBug
## 94                                the weather channel app
## 95                                                Bing
## 96                                   Weather Channel App
## 97                                    Talk to my mother.
## 98                                        i look outside
## 99                                Weather Channel App in IPhone
```

```
## 100                                                Weather Channel
## 101                                The weather app on the iPhone
## 102                                  The weather channel iPad app
## 103                                               weather underground
## 104 My husband usually informs me about the next day's weather.
## 105                                               Default iPhone app
## 106                                                       Google now
## 107                                                       Chrome app
## 108                                                Apple Weather App
## 109                                               Iphone weather app
## 110                                                     desktop icon
## 111                              The one that comes on your iphone
## 112                          the weather channel and local TV kxan
## 113                                                   Wunderground
## 114                       Intellicast (site) & WUnderground (app)
## 115                                                         myphone
## 116                                                  weather.gov
## 117                                my phone has a pre-loaded app
## 118                                               www.weather.gov
## 119                                     Weather & Clock widget
## 120                                          yahoo weather app
## 121                                            Www.weather.com
## 122                                           weather app on iphone
```

```
# Lowercase and trim values
weather <- weather |>
  mutate(
    `Other website or app` = tolower(trimws(`Other website or app`))
  )

# Standardize and correct values
weather <- weather |>
  mutate(`Other website or app` = case_when(
    `Other website or app` == "1 weather" ~ "1Weather",
    `Other website or app` == "1weather" ~ "1Weather",
    `Other website or app` == "accuweather" ~ "Accuweather",
    `Other website or app` == "accuweather app" ~ "Accuweather",
    `Other website or app` == "accuweather or weather underground" ~ "Multiple",
    `Other website or app` == "aol" ~ "Internet",
    `Other website or app` == "app on iphone" ~ "Default weather app",
    `Other website or app` == "apple weater" ~ "Default weather app" ,
    `Other website or app` == "apple weather app" ~ "Default weather app",
    `Other website or app` == "apple-provided site" ~ "Default weather app",
    `Other website or app` == "basic weather app on my iphone" ~ "Default weather app",
    `Other website or app` == "bing" ~ "Internet",
    `Other website or app` == "chrome app" ~ "Internet",
    `Other website or app` == "dark sky" ~ "Dark Sky",
    `Other website or app` == "default iphone app" ~ "Default weather app",
    `Other website or app` == "desktop icon" ~ "None",
    `Other website or app` == "directv weather app" ~ "DirecTV",
    `Other website or app` == "fancyclock app on my phone and accuweather.com" ~ "Multiple",
    `Other website or app` == "goes west satalite" ~ "GOES-West",
    `Other website or app` == "google \"weather\"" ~ "Internet",
    `Other website or app` == "google app" ~ "Internet",
```

```
    `Other website or app` == "google now" ~ "Internet",
    `Other website or app` == "i look outside" ~ "None",
    `Other website or app` == "intellicast" ~ "Weather Underground",
    `Other website or app` == "intellicast (site) & wunderground (app)" ~ "Weather Underground",
    `Other website or app` == "intellicast / storm" ~ "Weather Underground",
    `Other website or app` == "internet, tv, radio, newspaper" ~ "Multiple",
    `Other website or app` == "iphone" ~ "Default weather app",
    `Other website or app` == "iphone app" ~ "Default weather app",
    `Other website or app` == "iphone weather app" ~ "Default weather app",
    `Other website or app` == "ipod weather" ~ "Default weather app",
    `Other website or app` == "kcra online" ~ "Local TV Website",
    `Other website or app` == "local newspaper online  weather bug  noah" ~ "Multiple",
    `Other website or app` == "look/feel the weather outside" ~ "None",
    `Other website or app` == "m.accuweather.com" ~ "Accuweather",
    `Other website or app` == "my husband usually informs me about the next day's weather." ~ "None",
    `Other website or app` == "my phone has a pre-loaded app" ~ "Default weather app",
    `Other website or app` == "myphone" ~ "Default weather app",
    `Other website or app` == "national weather service" ~ "National Weather Service",
    `Other website or app` == "national weather service site" ~ "National Weather Service",
    `Other website or app` == "nice" ~ "None",
    `Other website or app` == "noaa and the weather channel" ~ "Multiple",
    `Other website or app` == "noaa app" ~ "NOAA",
    `Other website or app` == "noaa.gov" ~ "NOAA",
    `Other website or app` == "numerous local weather apps" ~ "Multiple",
    `Other website or app` == "phone" ~ "Default weather app",
    `Other website or app` == "smart phone" ~ "Default weather app",
    `Other website or app` == "storm team 4" ~ "Local TV Website",
    `Other website or app` == "talk to my mother." ~ "None",
    `Other website or app` == "the one that comes on your iphone" ~ "Default weather app",
    `Other website or app` == "the weather app on the iphone" ~ "Default weather app",
    `Other website or app` == "the weather channel" ~ "The Weather Channel",
    `Other website or app` == "the weather channel and local tv kxan" ~ "Multiple",
    `Other website or app` == "the weather channel app" ~ "The Weather Channel",
    `Other website or app` == "the weather channel app on my phone" ~ "The Weather Channel",
    `Other website or app` == "the weather channel app on my phone." ~ "The Weather Channel",
    `Other website or app` == "the weather channel for iphone or ipad" ~ "The Weather Channel",
    `Other website or app` == "the weather channel ipad app" ~ "The Weather Channel",
    `Other website or app` == "via my iphone builtin app" ~ "Default weather app",
    `Other website or app` == "weather & clock widget" ~ "Default weather app",
    `Other website or app` == "weather app" ~ "Default weather app",
    `Other website or app` == "weather app on iphone" ~ "Default weather app",
    `Other website or app` == "weather bug" ~ "WeatherBug",
    `Other website or app` == "weather bug app" ~ "WeatherBug",
    `Other website or app` == "weather channel" ~ "The Weather Channel",
    `Other website or app` == "weather channel app" ~ "The Weather Channel",
    `Other website or app` == "weather channel app android" ~ "The Weather Channel",
    `Other website or app` == "weather channel app in iphone" ~ "The Weather Channel",
    `Other website or app` == "weather channel app on my phone and ipad" ~ "The Weather Channel",
    `Other website or app` == "weather channel app on phone" ~ "The Weather Channel",
    `Other website or app` == "weather channel.com" ~ "The Weather Channel",
    `Other website or app` == "weather kitty" ~ "Weather Kitty",
    `Other website or app` == "weather puppy" ~ "Weather Puppy",
    `Other website or app` == "weather risk" ~ "Weather Risk",
```

```
    `Other website or app` == "weather timeline (android)" ~ "Weather Timeline",
    `Other website or app` == "weather underground" ~ "Weather Underground",
    `Other website or app` == "weather underground through hd widgets" ~ "Weather Underground",
    `Other website or app` == "weather underground, also local tv news half the time." ~ "Multiple",
    `Other website or app` == "weather.com" ~ "The Weather Channel",
    `Other website or app` == "weather.gov" ~ "National Weather Service",
    `Other website or app` == "weatherbug" ~ "WeatherBug",
    `Other website or app` == "weatherbug app" ~ "WeatherBug",
    `Other website or app` == "weatherbug on my phone" ~ "WeatherBug",
    `Other website or app` == "weatherbug, storm" ~ "Multiple",
    `Other website or app` == "weatherbug.com" ~ "WeatherBug",
    `Other website or app` == "weatherforyou.com" ~ "Weather For You",
    `Other website or app` == "weatherunderground" ~ "Weather Underground",
    `Other website or app` == "wunderground" ~ "Weather Underground",
    `Other website or app` == "wunderground.com" ~ "Weather Underground",
    `Other website or app` == "www.weather.com" ~ "The Weather Channel",
    `Other website or app` == "www.weather.gov" ~ "National Weather Service",
    `Other website or app` == "www.wunderground.com" ~ "Weather Underground",
    `Other website or app` == "yahoo" ~ "Yahoo Weather",
    `Other website or app` == "yahoo weather" ~ "Yahoo Weather",
    `Other website or app` == "yahoo weather app" ~ "Yahoo Weather",
    `Other website or app` == "yahoo weather iphone" ~ "Yahoo Weather",
    `Other website or app` == "yo window" ~ "YoWindow",
    `Other website or app` == "yr.no" ~ "yr.no",
    .default = `Other website or app`
  ))
```

After standardizing the free response values, we corrected the data for when users should have chosen a different response instead of Specific Website of App in the Method of checking weather column. We changed the respondent's response for the Method of checking weather column and removed their response from the Other website or app column, as they wouldn't have given a response if they chose the correct option. We also noticed that some of the free responses didn't mention any apps or websites, such as "talk to my mother.". We chose to replace these answers to None and set their Checks daily weather report data to False because we wanted to focus on whether respondents used any apps or websites to check the weather.

```
# Correct Method of checking weather columns based on other column
weather <- weather |>
  mutate(`Method of checking weather` =
    case_when(
      `Other website or app` %in% methods ~ `Other website or app`,
      .default = `Method of checking weather`
    ),
    `Other website or app` =
      case_when(
        `Other website or app` %in% methods ~ "-",
        .default = `Other website or app`
      )
  )


# Change checks daily weather if doesn't use any apps/websites to check
weather <- weather |>
  mutate(`Checks daily weather` =
           case_when(
```

```
            `Other website or app` == "None" ~ as.factor(FALSE),
            .default = `Checks daily weather`
          ),
        `Other website or app` =
          case_when(
            `Other website or app` == "None" ~ "-",
            .default = `Other website or app`
          )
      )
```

Then we replace the hyphen symbols with actual missing values so that our analysis doesn't mistaken a hyphen value as a valid separate category.

```
# Replace - characters with missing value
weather <- weather |>
  mutate(
    across(where(is.character), ~na_if(., "-"))
  )
```

All of the other columns were already fairly clean aside from using the hyphen as a placeholder for missing values. For each of the other columns, we converted the values into factors and set a order if there was some natural ordering to the categories.

For the Check weather on smartwatch column, all of the data were based on a scale of likely to unlikely. We ordered the values by most likely to least likely.

```
# Factor and reorder smart watch column
weather |>
  select(`Check weather on smartwatch`) |>
  distinct()
```

```
##   Check weather on smartwatch
## 1                 Very likely
## 2             Somewhat likely
## 3               Very unlikely
## 4                        <NA>
## 5           Somewhat unlikely
```

```
weather <- weather |>
  mutate(`Check weather on smartwatch` = factor(`Check weather on smartwatch`,
                                                levels = c("Very likely", "Somewhat likely", "Somewha
```

Then for the age column, we ordered the values from youngest age range to oldest.

```
# Factor and reorder age
weather |>
  select(Age) |>
  distinct()
```

```
##       Age
## 1 30 - 44
## 2 18 - 29
## 3    <NA>
## 4 45 - 59
## 5     60+
```

```
weather <- weather |>
  mutate(Age = factor(Age, levels = c("18 - 29", "30 - 44", "45 - 59", "60+")))
```

For the gender column, we ordered based on alphabetical order.

```
# Factor and reorder gender
weather |>
  select(Gender) |>
  distinct()
```

```
##   Gender
## 1   Male
## 2   <NA>
## 3 Female
```

```
weather <- weather |>
  mutate(Gender = factor(Gender, levels = c("Female", "Male")))
```

In the next column, household income, we ordered the values by lowest range to highest range and then prefer not to answer.

```
# Factor and reorder household income
weather |>
  select(`Household income`) |>
  distinct()
```

```
##        Household income
## 1    $50,000 to $74,999
## 2  Prefer not to answer
## 3  $100,000 to $124,999
## 4  $150,000 to $174,999
## 5    $25,000 to $49,999
## 6                  <NA>
## 7          $0 to $9,999
## 8    $10,000 to $24,999
## 9    $75,000 to $99,999
## 10     $200,000 and up
## 11 $175,000 to $199,999
## 12 $125,000 to $149,999
```

```
income_levels <- c("$0 to $9,999", "$10,000 to $24,999", "$25,000 to $49,999", "$50,000 to $74,999", "$

weather <- weather |>
  mutate(
    `Household income` = factor(`Household income`, levels = income_levels)
  )
```

Lastly, we turned region into factors with no specific ordering since there was no natural ordering of these values.

```
# Factor regions
weather |>
  select(`Region`) |>
  distinct()
```

```
##                Region
## 1      South Atlantic
## 2                <NA>
## 3     Middle Atlantic
## 4  West South Central
## 5             Pacific
```

```
## 6  West North Central
## 7  East North Central
## 8            Mountain
## 9         New England
## 10 East South Central
```

```
weather <- weather |>
  mutate(Region = factor(Region))
```

After wrangling all the data, we take a look at the resulting data set. The columns are all more readable and concise and the values are all standardized across each column, including the free response column. We also fixed some values based on user errors when a user should have selected a different response instead. The resulting data set is now cleaned and ready for analysis.

```
# Review wrangled data
glimpse(weather)
```

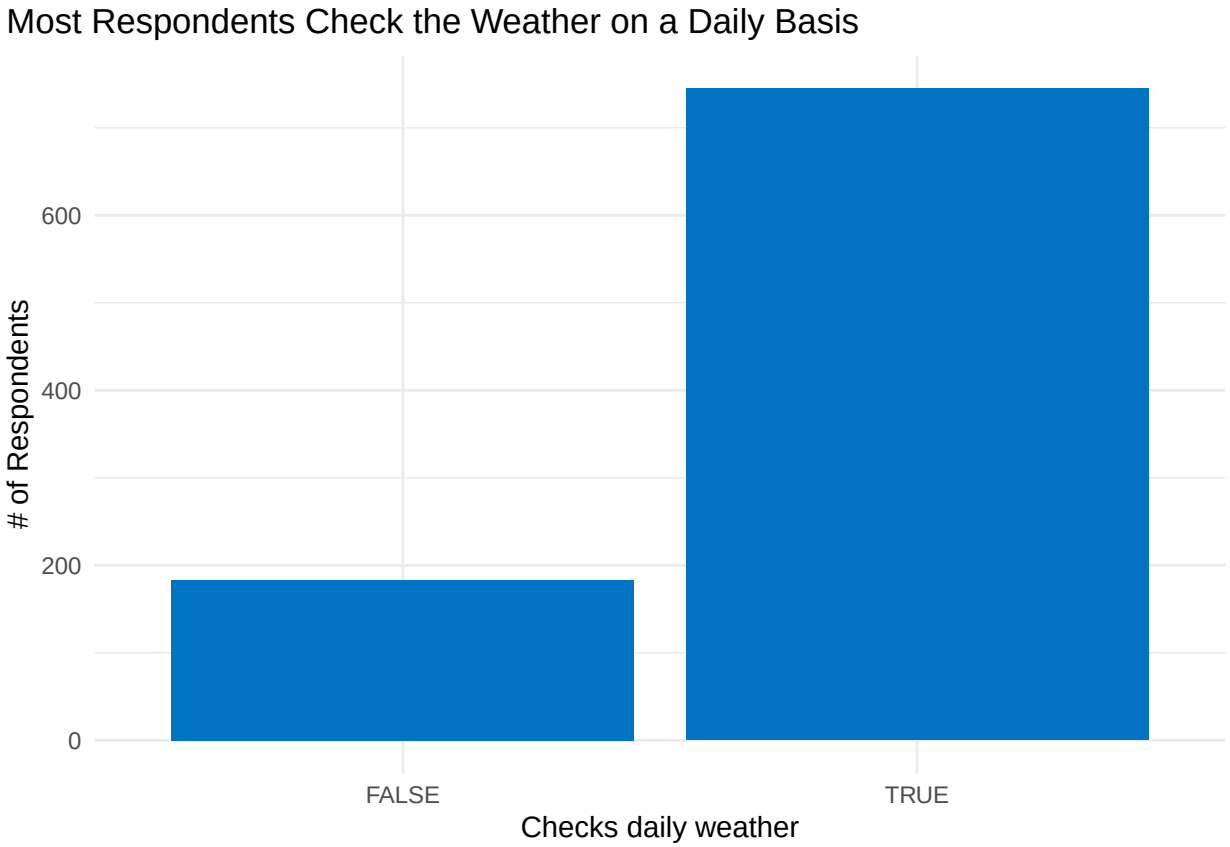
```
## Rows: 928
## Columns: 9
## $ id                              <dbl> 3887201482, 3887159451, 3887152228, 3887~
## $ `Checks daily weather`          <fct> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE~
## $ `Method of checking weather`    <chr> "Default weather app", "Default weather ~
## $ `Other website or app`          <chr> NA, NA, NA, NA, NA, "Accuweather", NA, N~
## $ `Check weather on smartwatch`   <fct> Very likely, Very likely, Very likely, S~
## $ Age                             <fct> 30 - 44, 18 - 29, 30 - 44, 30 - 44, 30 -~
## $ Gender                          <fct> Male, Male, Male, Male, Male, Male, Male~
## $ `Household income`              <fct> "$50,000 to $74,999", "Prefer not to ans~
## $ Region                          <fct> South Atlantic, NA, Middle Atlantic, NA,~
```

Analysis

Exploratory Data Analysis

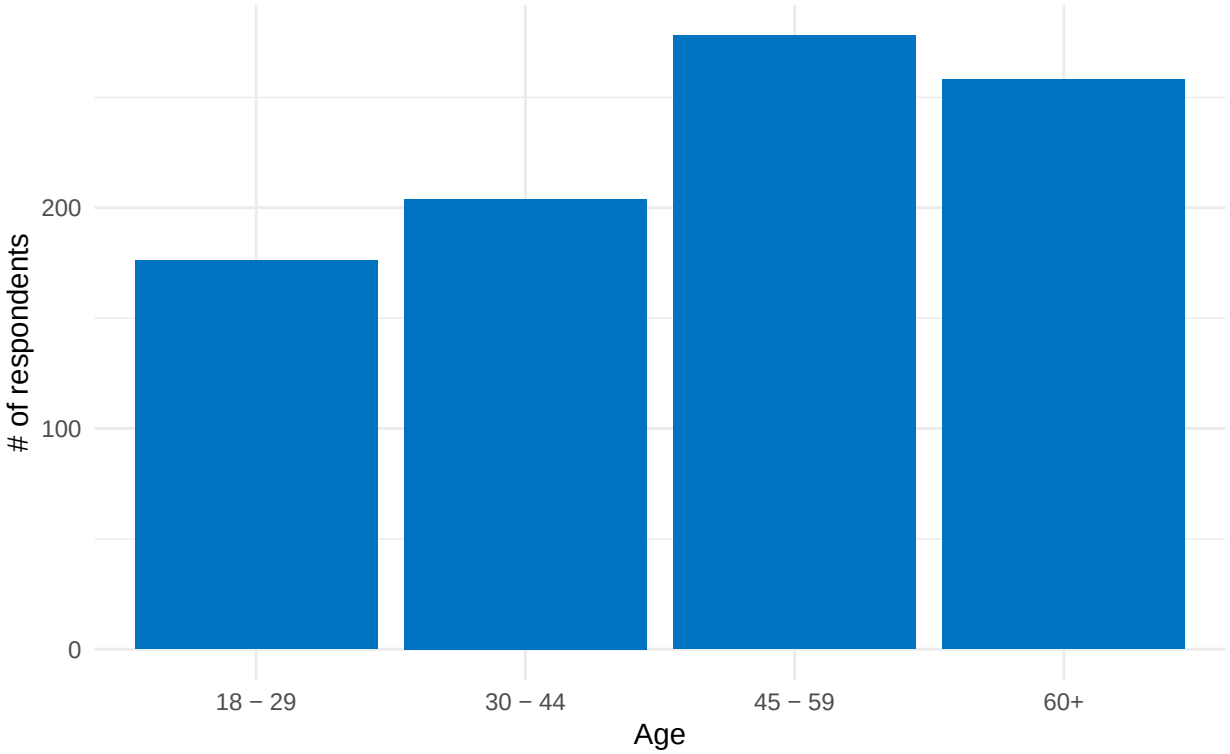
After wrangling our data, we explored the distributions of our data. We first plotted the number of survey responses that indicated whether they typically checked a daily weather report. From our plot, we can see that an overwhelming majority of respondents do typically check a daily weather report, while nearly 20% of people don't.

```
ggplot(weather, aes(`Checks daily weather`)) +
  geom_bar(fill = "#0073C2FF") +
  labs(title = "Most Respondents Check the Weather on a Daily Basis",
       y = "# of Respondents") +
  theme_minimal() +
  theme(plot.title.position = "plot")
```



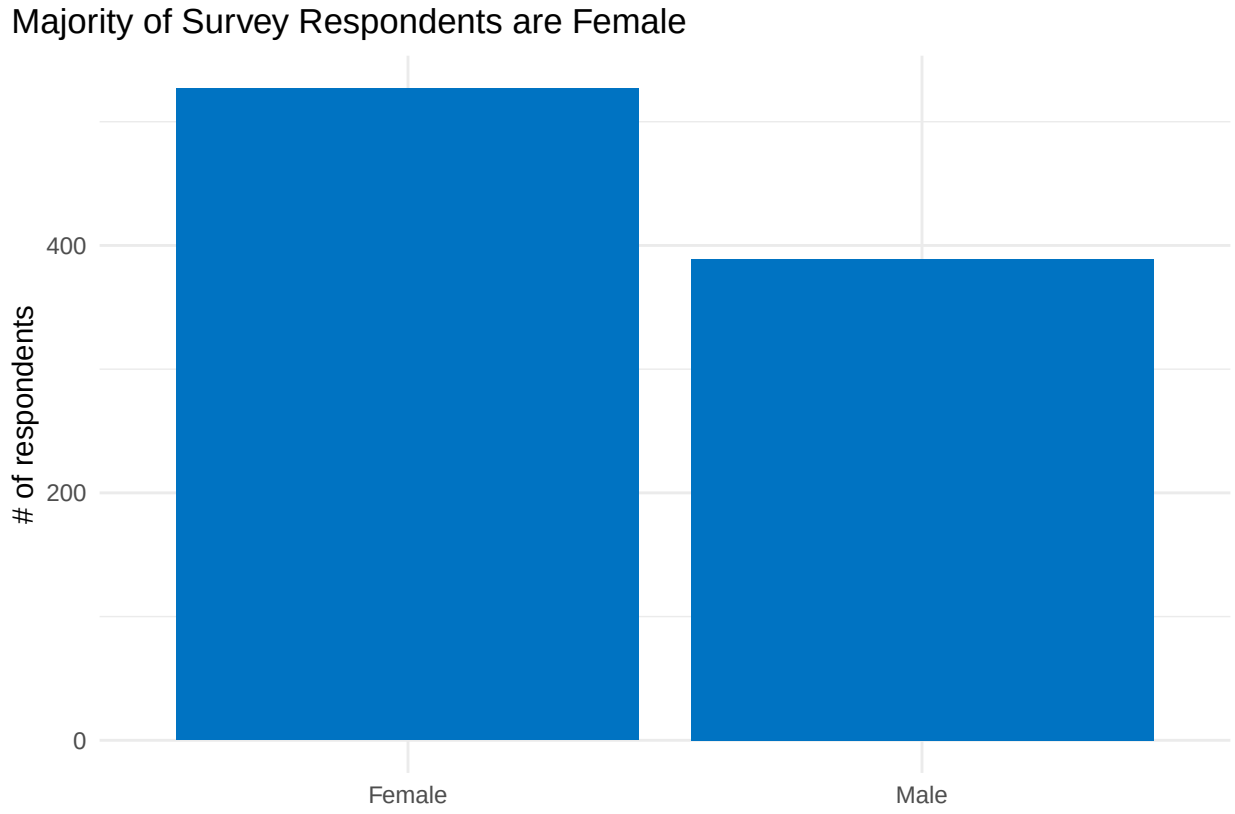
Next, we explored the age distribution of the respondents. Most of the respondents come from 45-59 year range and the 60+ years range. There are less respondents younger than 44 years old than there are those above 44 years old. However, we still have a large number of respondents from each age group in our data set.

```
age_dist <- weather |>
  filter(!is.na(Age))

ggplot(age_dist, aes(Age)) +
  geom_bar(fill = "#0073C2FF") +
  labs(title = "Age of Survey Respondents",
       subtitle = "Most of the observations in this data set come from those above the age of 45",
       y = "# of respondents") +
  theme_minimal() +
  theme(plot.title.position = "plot")
```



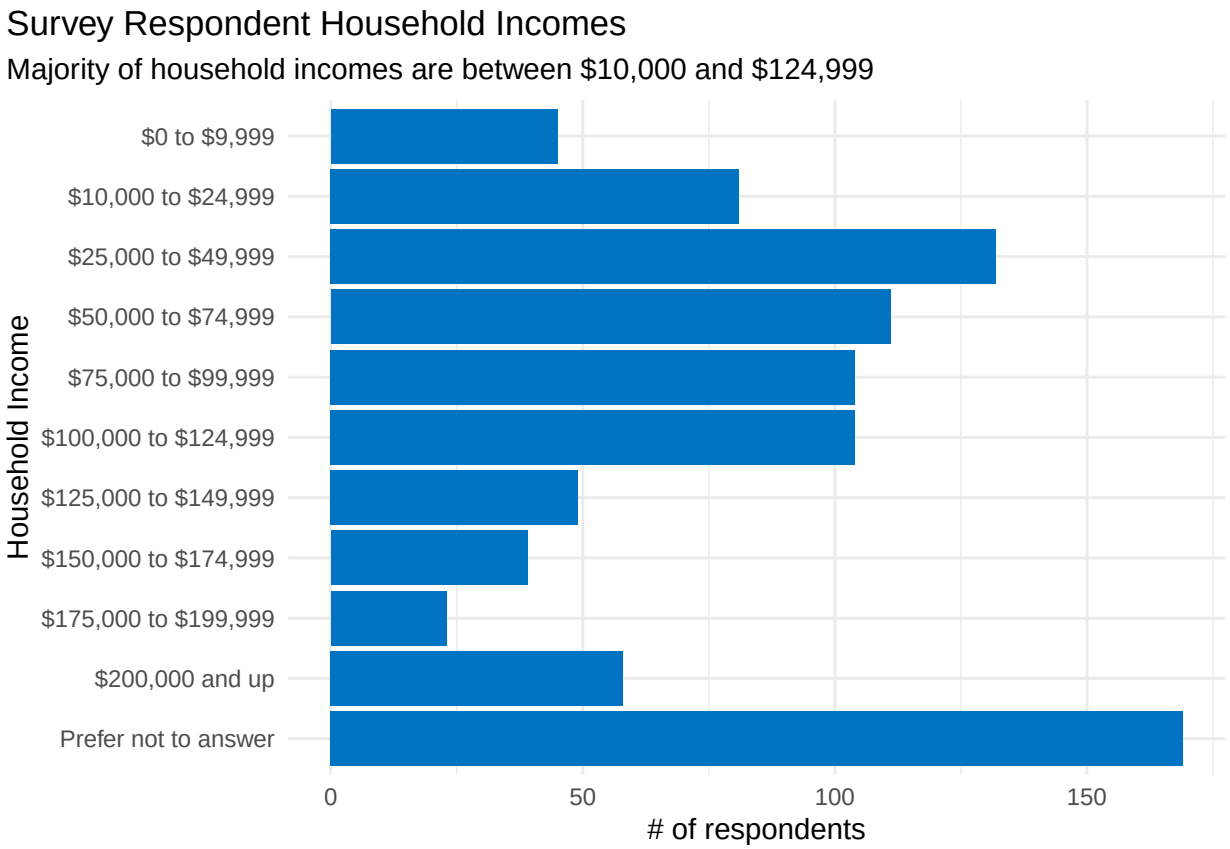
We then explored the distribution of gender across our data set. From our plot, we can see that a majority of survey respondents are female, with the male respondents differing by more than 100 respondents.

```
gender_dist <- weather |>
  filter(!is.na(Gender))

ggplot(gender_dist, aes(Gender)) +
  geom_bar(fill = "#0073C2FF") +
  labs(
    title = "Majority of Survey Respondents are Female",
    x = "",
    y = "# of respondents"
  ) +
  theme_minimal() +
  theme(plot.title.position = "plot")
```



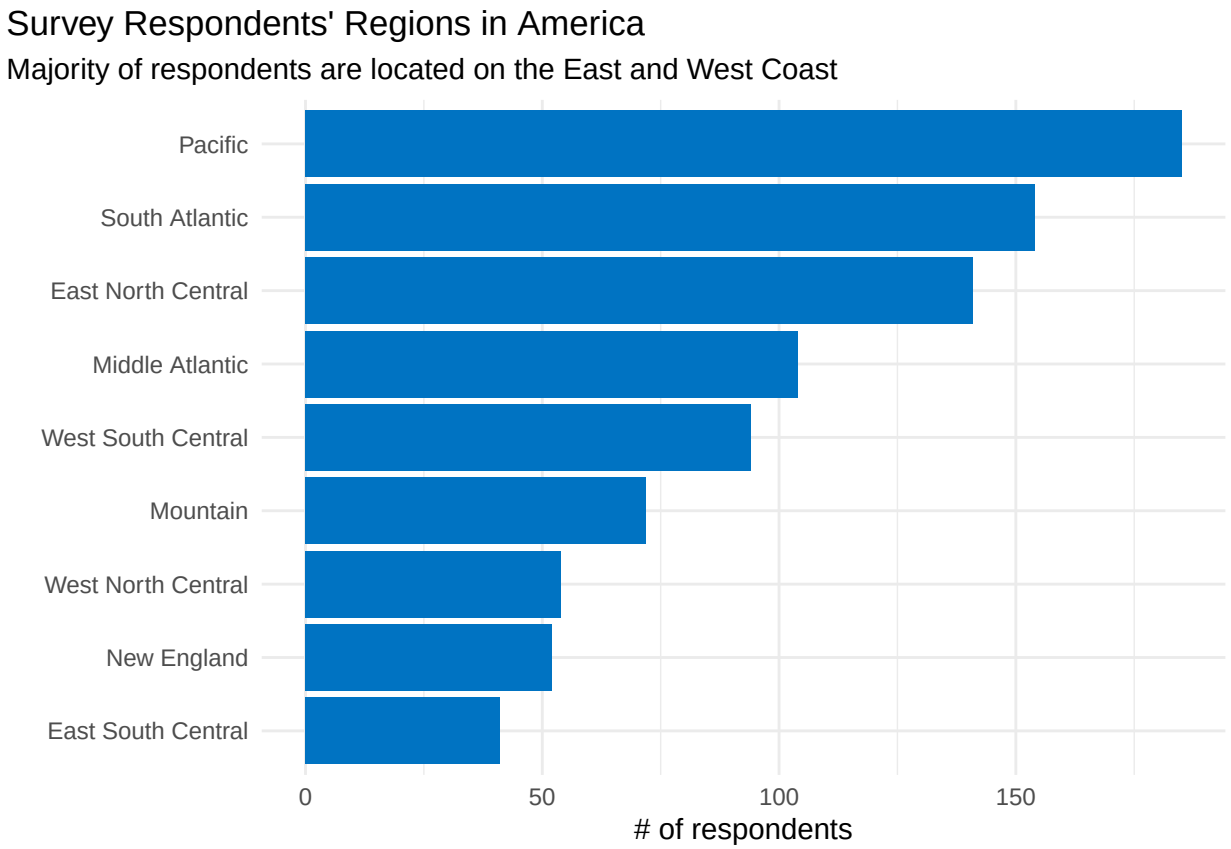
We also explored the distribution of household incomes in our data set. While there are a lot of respondents who chose not to include their household income in this survey, we do have a large majority of respondents who did select a household income range. A majority of the respondents who included their household income fell within the $10,000 to $124,999 range, with a large proportion of respondents within the $25,000 to $49,999 range. Interestingly, we have quite a few number of outliers in the $200,000 and up group.

```
income_dist <- weather |>
  filter(!is.na(`Household income`))

ggplot(income_dist, aes(fct_rev(`Household income`))) +
  geom_bar(fill = "#0073C2FF") +
  coord_flip() +
  labs(title = "Survey Respondent Household Incomes",
       subtitle = "Majority of household incomes are between $10,000 and $124,999",
       x = "Household Income",
       y = "# of respondents") +
  theme_minimal() +
  theme(plot.title.position = "plot")
```



Then, we explored the distribution of regions of America the respondents were from. We noticed that there were a large number of respondents from the East and West Coast regions, while the central regions had relatively less number of respondents. Additionally, the Pacific region has the most number of respondents to this survey.

```
region_dist <- weather |>
  filter(!is.na(Region))

region_dist |>
  group_by(Region) |>
  summarise(counts = n()) |>
  ggplot(aes(x = reorder(Region, counts), y = counts)) +
  geom_bar(stat="identity", fill = "#0073C2FF") +
  labs(title = "Survey Respondents' Regions in America",
       subtitle = "Majority of respondents are located on the East and West Coast",
       x = "",
       y = "# of respondents") +
  coord_flip() +
  theme_minimal() +
  theme(plot.title.position = "plot")
```



Data Analysis

After exploring our data, we investigated the questions that we proposed from before. We used plots and logistic regression models to analyze our data to answer these questions. Do note that the following logistic regressions will not be accounting for survey weighting, since this information was not included in the data set.

The first question we sought to answer was:

> Is there a relationship between whether a person typically checks their weather app and their sex?

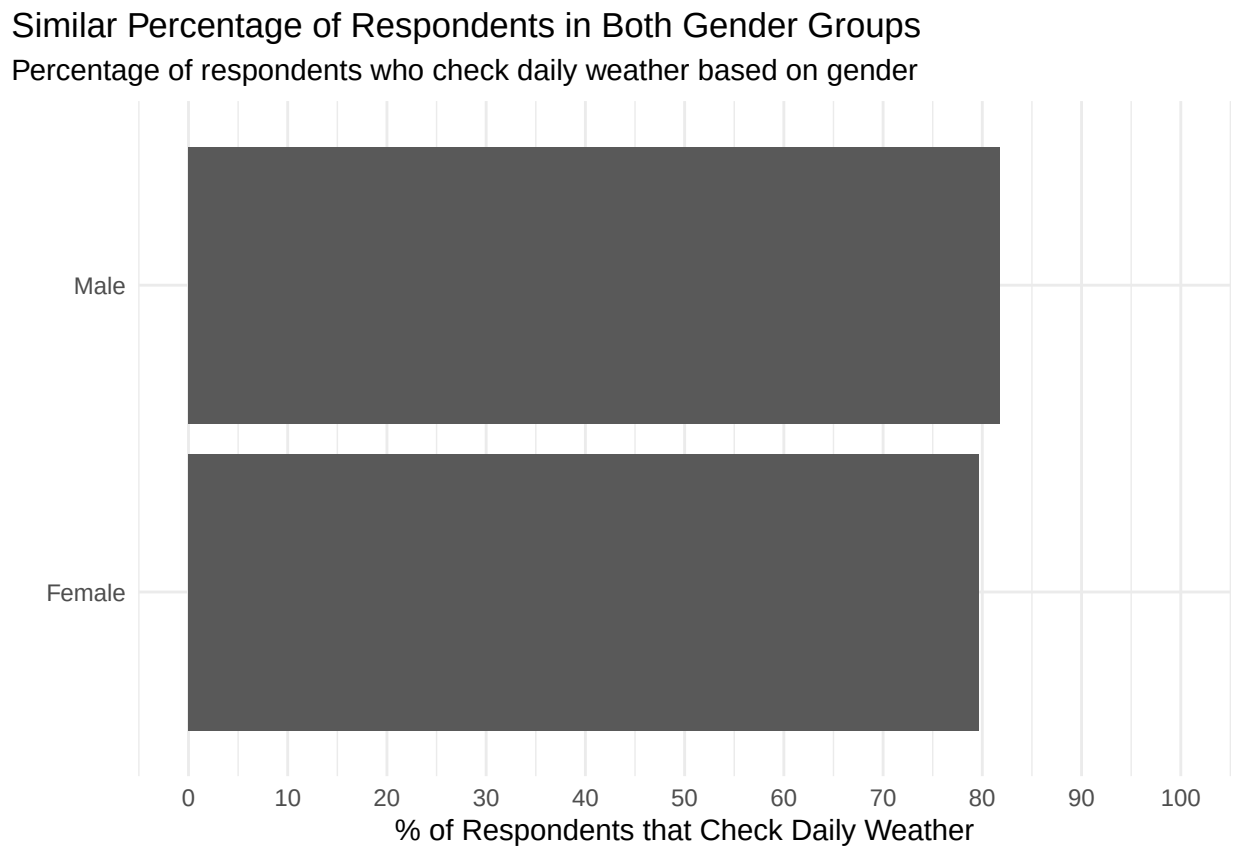
We began with plotting the percentage of those who typically check a daily weather report based on their sex. We created a table and plot displaying the percentages to visualize the difference in both groups' distributions. While there were much more females responding in this survey, both groups had very similar percentage of those who typically check a daily weather report.

```
weather_gender <- weather |>
  filter(!is.na(Gender)) |>
  group_by(Gender) |>
  summarize(`Checks Daily Weather` = mean(as.logical(`Checks daily weather`) * 100)) |>
  mutate("Doesn't Check Daily Weather" = 100 - `Checks Daily Weather`) |>
  arrange(desc(`Checks Daily Weather`))

# Display percentages
weather_gender |>
  kbl(caption = "Percentage of Users in Different Gender Groups who Check Daily Weather") |>
  kable_material(c("striped", "hover"))
```

Table 1: Percentage of Users in Different Gender Groups who Check Daily Weather

Gender	Checks Daily Weather	Doesn't Check Daily Weather
Male	81.74807	18.25193
Female	79.69639	20.30361

```
# Plot percentage of users who check daily weather by gender
ggplot(weather_gender, aes(x = `Checks Daily Weather`, y = Gender)) +
  geom_bar(stat = 'identity') +
  scale_x_continuous(breaks = seq(0, 100, by = 10),
                     labels = seq(0, 100, by = 10),
                     limits = c(0, 100)) +
  labs(
    title = "Similar Percentage of Respondents in Both Gender Groups",
    subtitle = "Percentage of respondents who check daily weather based on gender",
    y = "",
    x = "% of Respondents that Check Daily Weather"
  ) +
  theme_minimal() +
  theme(
    plot.title.position = "plot"
  )
```

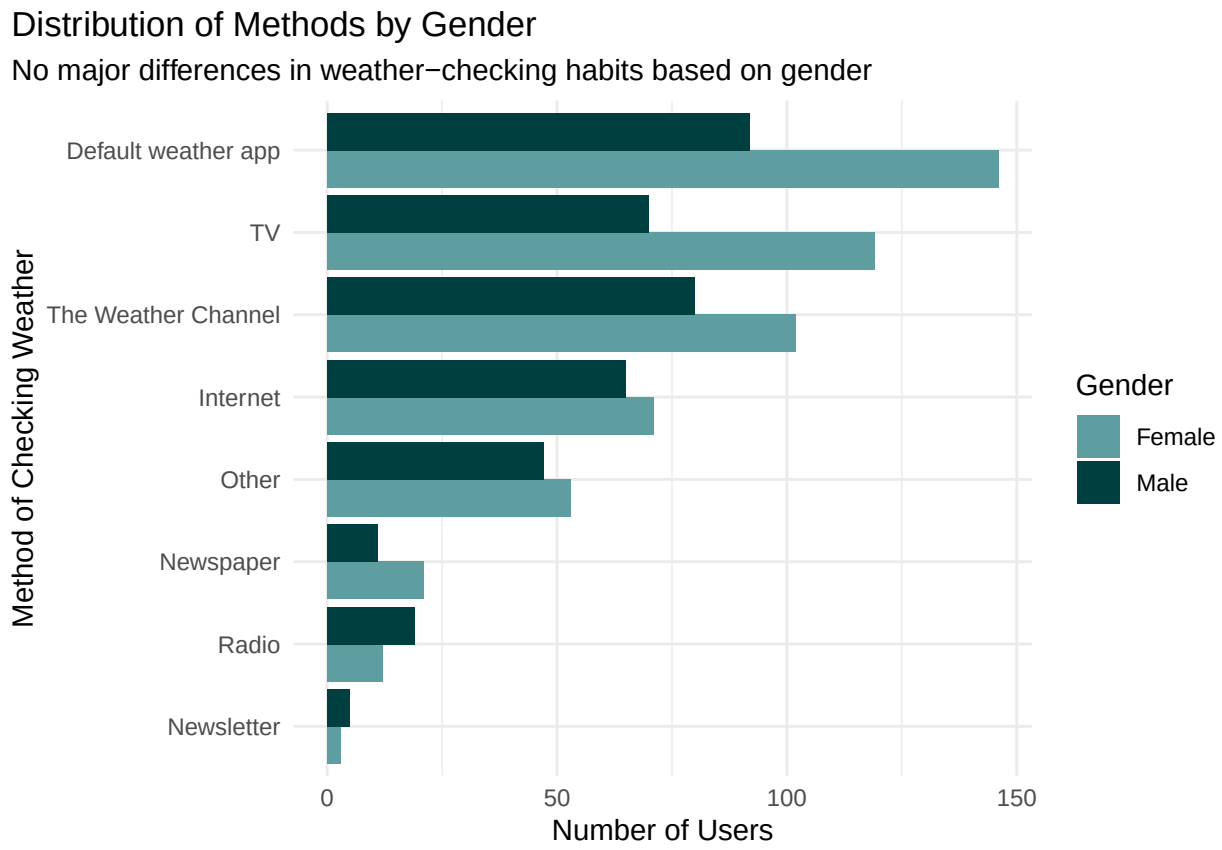


Next, we also explored if the method of checking the weather report differed between sexes. We created a plot to visualize the number of male and female respondents and what method they use to check the weather.

Overall, there doesn't seem to be any major differences in the relative distributions across the different methods of checking the weather.

```
weather |>
  filter(is.na(Gender) == FALSE) |>
  group_by(Gender, `Method of checking weather`) |>
  summarize(group_count = n()) |>
  mutate(`Method of checking weather` = fct_reorder(`Method of checking weather`, group_count)) |>
  ggplot(aes(fill=Gender, y=`Method of checking weather`, x=group_count)) +
         geom_bar(position = "dodge", stat="identity") +
  labs(title = "Distribution of Methods by Gender",
       subtitle = "No major differences in weather-checking habits based on gender",
       x = "Number of Users",
       y= "Method of Checking Weather") +
  theme_minimal() +
  theme(plot.title.position = "plot") +
  scale_fill_manual(values = c("cadetblue", "#003f3f"))
```

```
## `summarise()` has grouped output by 'Gender'. You can override using the
## `.groups` argument.
```



We also wanted to quantify the likelihood that a respondent typically checks a daily weather report based on their gender. We displayed the estimated coefficients in the table, and get the following logisitic regression model:

$\log(p/(1\text{-}p)) = 1.367 + 0.131$ (Male)

The resulting model tells us that the log likelihood of a female respondent typically checking a daily weather

Table 2: Coefficient Estimates of Logistic Regression Model Using Gender

term	estimate
(Intercept)	1.3674259
GenderMale	0.1319456

report is 1.367. On the other hand, if the respondent was male, then the likelihood would increase by 0.131 to a log likelihood of 1.498. Thus, from our dataset, males are more likely to typically check the daily weather report compared to females.

```
# Create log reg model predicting whether respondent uses daily weather report based on gender
lr_gender <- logistic_reg() |>
  set_engine("glm") |>
  fit(`Checks daily weather` ~ Gender, data=weather, family = "polynomial")

lr_gender |>
  tidy() |>
  select(term, estimate) |>
  kbl(caption = "Coefficient Estimates of Logistic Regression Model Using Gender") |>
  kable_material(c("striped", "hover"))
```

The next question we wanted to answer was

> Does weather app use differ based on where the respondent is located?

To answer this question, we first investigated the percentage of users across regions who typically check and don't check the daily weather report. We created a table for each region and the percentage of those who did indicate that they typically checked the daily weather report versus does who didn't. A large majority of regions have over 80% of respondents who indicated that they typically check the daily weather report. There doesn't seem to be any clear correlation between general geographic proximity and the percentage of users as region with less than 80% rate are regions scattered across America. However, it seems that most of the central regions of America tend to have a high percentage of respondents who typically check the daily weather report, whereas some of the coastal regions, like South Atlantic and Pacific, have a smaller percentage. Interestingly, respondents who didn't include what region they were from generally had a higher rate of not checking the daily weather report, at 45%, which is the highest compared to the other regions.

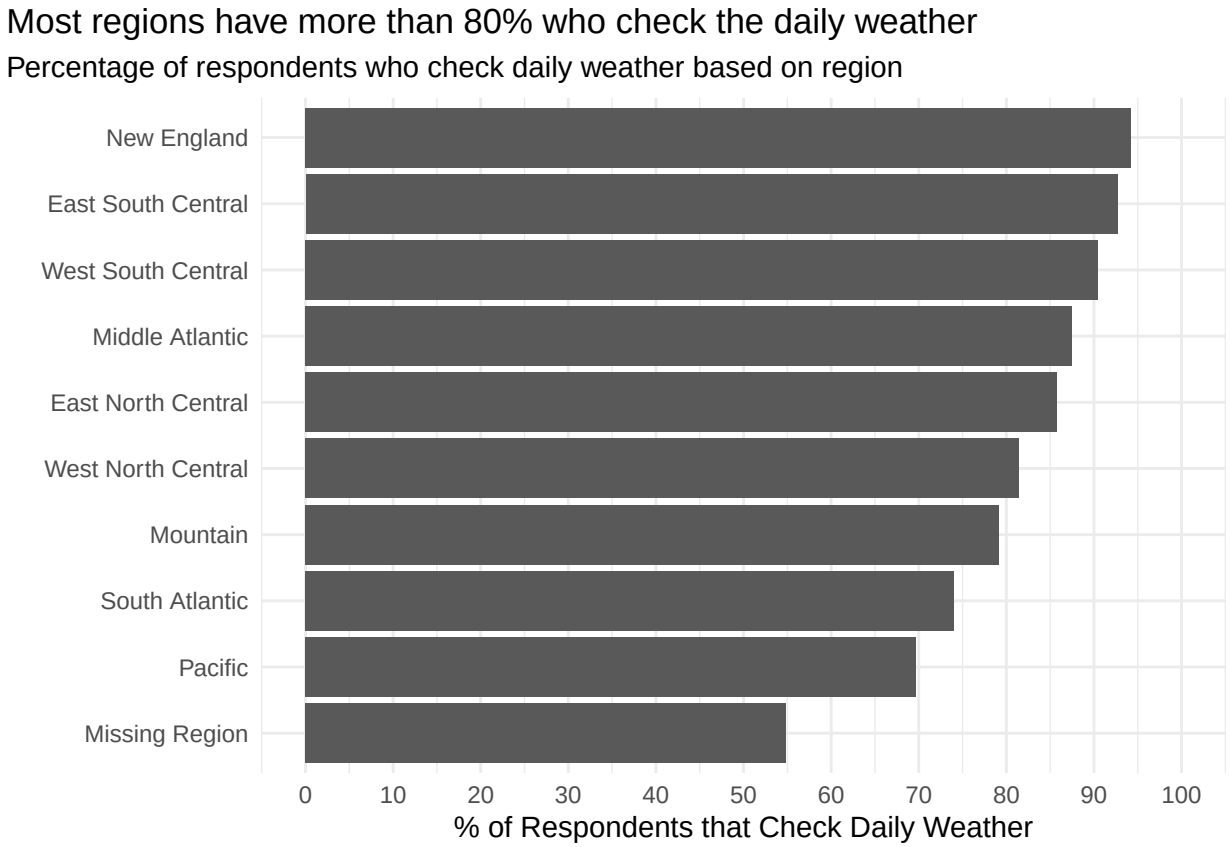
```
# Table of percentage of users who do and don't check weather app by region
weather_region <- weather |>
  mutate(Region = as.character(Region)) |>
  group_by(Region) |>
  summarize(`Checks Daily Weather` = mean(as.logical(`Checks daily weather`) * 100)) |>
  mutate("Doesn't Check Daily Weather" = 100 - `Checks Daily Weather`) |>
  arrange(desc(`Checks Daily Weather`)) |>
  replace_na(list(Region = "Missing Region"))

weather_region |>
  kbl(caption = "Percentage of Users Across Regions who Check Daily Weather") |>
  kable_material(c("striped", "hover"))
```

Then we plotted the percentages of those who did typically check the daily weather report by region to visualize the differences. The New England region has the highest percentage of 94%, followed by East South Central with 92%, and West South Central with 90%. Again, we see that most of the central regions tend to have a fairly high percentage of respondents who check the daily weather report compared to all the coastal regions. Only the New England and Middle Atlantic regions deviate from this pattern, as they make up the top four regions with the highest percentage of respondents who typically check the daily weather report.

Table 3: Percentage of Users Across Regions who Check Daily Weather

Region	Checks Daily Weather	Doesn't Check Daily Weather
New England	94.23077	5.769231
East South Central	92.68293	7.317073
West South Central	90.42553	9.574468
Middle Atlantic	87.50000	12.500000
East North Central	85.81560	14.184397
West North Central	81.48148	18.518518
Mountain	79.16667	20.833333
South Atlantic	74.02597	25.974026
Pacific	69.72973	30.270270
Missing Region	54.83871	45.161290

```
# Plot percentage of users who check daily weather by region
ggplot(weather_region, aes(x = `Checks Daily Weather`, y = fct_reorder(Region, `Checks Daily Weather`)))
  geom_bar(stat = 'identity') +
  scale_x_continuous(breaks = seq(0, 100, by = 10),
                     labels = seq(0, 100, by = 10),
                     limits = c(0, 100)) +
  labs(
    title = "Most regions have more than 80% who check the daily weather",
    subtitle = "Percentage of respondents who check daily weather based on region",
    y = "",
    x = "% of Respondents that Check Daily Weather"
  ) +
  theme_minimal() +
  theme(
    plot.title.position = "plot"
  )
```



While there doesn't seem to be any clear correlation between regions and whether a respondent checks the daily weather report, we wanted to quantify the likelihood of a respondent checking the daily weather report based on their region. This logistic regression doesn't consider any survey weights and might not be representative of the population of America, however, the model gives us some insight on how region affects the likelihood that a respondent from our sample typically checks the daily weather report.

We get the following logistic regression model predicting whether a user typically checks a daily weather report given their region. Note that the base line is the New England region, which had the highest percentage of respondents who typically checked a daily weather report.

log(p/(1-p)) = 2.793 - 0.993 (East North Central) - 0.254 (East South Central) + -0.847 (Middle Atlantic) - 1.458 (Mountain) - 1.958 (Pacific) - 1.745 (South Atlantic) - 1.311 (West North Central) - 0.547 (West South Central)

The base log likelihood value when the region is New England is equal to 2.793. Notice that all the other coefficients of other regions are all negative. This means that the likelihood of a person typically checking a daily weather report would be lower in any other region of America compared to someone from the New England region. The East South Central region has the next highest log likelihood value, with a coefficient of -0.254, which is the smallest magnitude coefficient. On the other hand, people from the Pacific region has the lowest log likelihood, with a coefficient of -1.958, or the largest negative coefficient of the model. The results we get from these models are very similar to the percentages we plotted per region, however, we quantified the likelihood across different regions and how they compare with other region likelihoods.

```
# Logistic regression based on region and whether person checks daily weather
relevel_region <- weather |>
  mutate(Region = relevel(`Region`, "New England"))
lr_region <- logistic_reg() |>
  set_engine("glm") |>
```

Table 4: Coefficient Estimates of Logistic Regression by Region

term	estimate
(Intercept)	2.7932080
East North Central	-0.9931497
East South Central	-0.2542341
Middle Atlantic	-0.8472979
Mountain	-1.4582069
Pacific	-1.9587473
South Atlantic	-1.7458890
West North Central	-1.3116035
West South Central	-0.5477813

```
  fit(`Checks daily weather` ~ Region, data = relevel_region, family = "binomial")

# Note: East North Central is base level
lr_region |>
  tidy() |>
  select(term, estimate) |>
  mutate(
    term = str_remove(term, "Region")
  ) |>
  kbl(caption = "Coefficient Estimates of Logistic Regression by Region") |>
  kable_material(c("striped", "hover"))
```

While many of the regions had a high percentage of people who typically check the daily weather report, we also wanted to see if the method in which they checked the daily weather report differed by region. We first created a table containing the percentage of people within each region that indicated that they used that method to check the weather report.

We noticed that regions do differ in how they check their daily weather report. In North East Central, the most popular method was the default weather app, making up 29%, followed by TV with 23%. For East South Central, the most popular method was TV with 31%, followed by Internet with 24%. In the Middle Atlantic region, the most popular method of checking the weather was TV with 24%, followed closely by The Weather Channel with 23%. In the Mountain region, the default weather app was the most popular method with 27%, followed by The Weather Channel with 23%. In New England, the default weather app was also the most popular method with 30%, followed by The Weather Channel with 21%. For the Pacific region, the most popular method was the default weather app with 23%, followed closely by Internet with 21%. In the South Atlantic region, the default weather app was the most popular by far, with 33%, with The Weather Channel being the next popular method with 20%. For the West North Central region, the default weather app was also the most popular method at 31%, followed by both The Weather Channel and TV at 20% each. Lastly, in the West South Central region, TV was the most popular method at 30%, followed by the default weather app and The Weather Channel, at 20% and 19% respectively. For the responses that didn't indicate any region, 35% also did not indicate a method, making up the majority of missing region responses.

```
# Create table of percentage of users based on method of checking and region

# Count number of users per method
method_region <- weather |>
  group_by(`Method of checking weather`, Region) |>
  summarize("count" = n()) |>
  mutate(Region = as.character(Region)) |>
  replace_na(list(Region = "Missing Region"))
```

Table 5: Percentage of Respondents' Method of Checking Weather

Region	Default weather app	Internet	Newsletter	Newspaper	Other	Radio	TV
East North Central	29.078014	9.929078	0.0000000	2.836879	13.475177	1.418440	23.404255
East South Central	7.317073	24.390244	0.0000000	0.000000	7.317073	7.317073	31.707317
Middle Atlantic	21.153846	8.653846	0.0000000	2.884615	13.461538	6.730769	24.038462
Mountain	27.777778	13.888889	0.0000000	4.166667	11.111111	1.388889	18.055556
New England	30.769231	11.538462	0.0000000	5.769231	9.615385	1.923077	19.230769
Pacific	23.783784	21.621622	1.6216216	4.864865	10.810811	3.783784	16.216216
South Atlantic	33.766234	11.688312	0.6493506	3.896104	11.038961	3.246753	15.584416
West North Central	31.481482	12.962963	0.0000000	3.703704	9.259259	1.851852	20.370370
West South Central	20.212766	17.021277	2.1276596	0.000000	7.446809	3.191489	30.851064
Missing Region	12.903226	19.354839	6.4516129	6.451613	6.451613	3.225807	3.225807

```
## `summarise()` has grouped output by 'Method of checking weather'. You can
## override using the `.groups` argument.
```

```
# Count number of respondents per region
region_cnt <- weather |>
  mutate(Region = as.character(Region)) |>
  replace_na(list(Region = "Missing Region")) |>
  group_by(Region) |>
  summarize("total" = n())

# Convert counts to percentages
method_region <- method_region |>
  left_join(region_cnt, by = c("Region")) |>
  mutate("percentage" = (count / total) * 100)

# Display table
wide_method_region <- method_region |>
  select(Region, percentage) |>
  pivot_wider(names_from = `Method of checking weather`, values_from = percentage) |>
  mutate_if(is.numeric, ~replace_na(., 0))
```

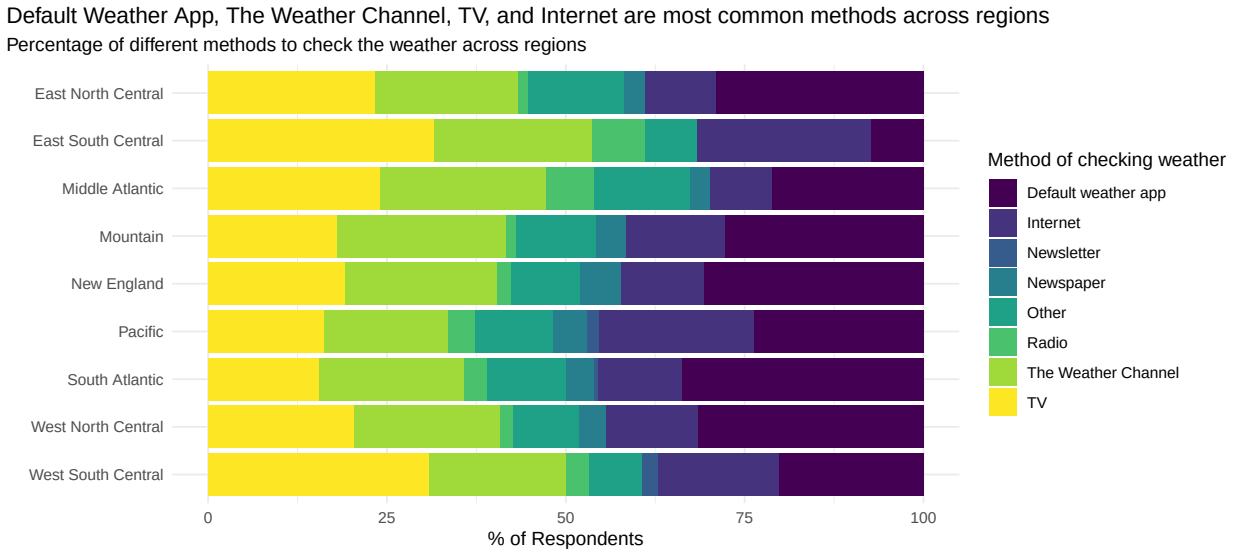
```
## Adding missing grouping variables: `Method of checking weather`
```

```
wide_method_region |>
  kbl(caption = "Percentage of Respondents' Method of Checking Weather") |>
  kable_material(c("striped", "hover"))
```

We then plotted the percentages of each method within each region. Similar to the results of the table we created, the percentages of each method and the most popular methods of checking the weather differ greatly per region. However, we do see that the default weather app, The Weather Channel, TV, and the Internet are the most popular methods across all regions.

```
# Plot different methods by region
region_order = wide_method_region |> pull(Region)
ggplot(method_region |>
         filter(!is.na(`Method of checking weather`), Region != "Missing Region") |>
         mutate(
           Region = factor(Region, levels = region_order)
         ), aes(x = percentage, y = fct_rev(Region), fill = `Method of checking weather`)) +
  geom_bar(stat = 'identity') +
```

```
  labs(
    title = "Default Weather App, The Weather Channel, TV, and Internet are most common methods across r
    subtitle = "Percentage of different methods to check the weather across regions",
    x = "% of Respondents",
    y = ""
  ) +
  theme_minimal() +
  theme(
    plot.title.position = "plot"
  ) +
  scale_fill_viridis(discrete = TRUE)
```



Next, we wanted to answer

> Is there a relationship between household income and weather app?

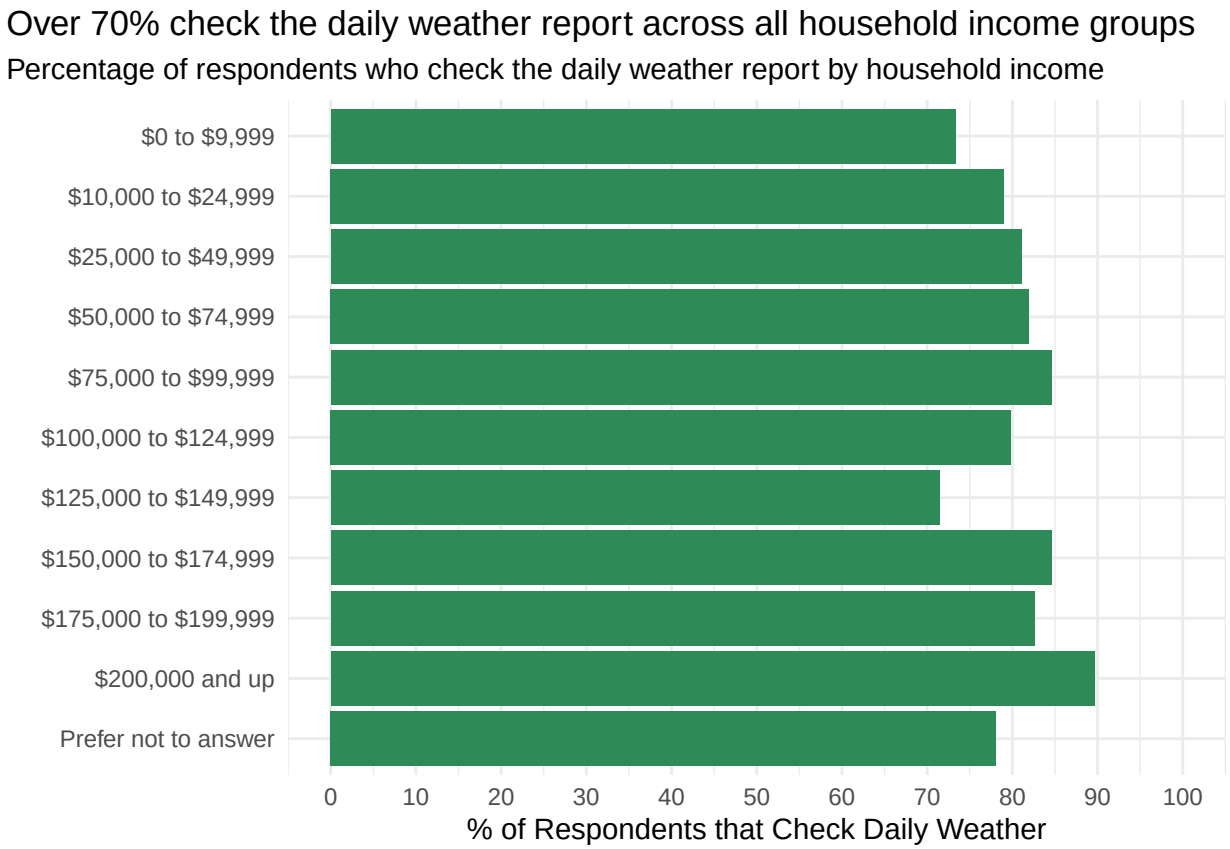
We first started with creating a table and plotting the percentage of respondents who typically check the daily weather report within each household income groups. Since the number of respondents from each group was unbalanced, viewing the percentage within each groups gives us a better comparison of how daily weather report use differs across each income group. We notice that across all income groups, more than 70% of respondents said that they check the daily weather report. The $200,000 and up group had the highest percentage of people within their group to respond that they do typically check the daily weather report with 89%. The lowest percentage group was $125,000 to $149,999 with only 71% of respondents within that group indicating that they typically check the daily weather report. The bar plot also shows that while there is some variation in the percentage across different income groups, the groups actually don't vary by much and have fairly high rates of respondents who check the daily weather report.

```
# Table of percentage of users who do and don't check weather app by household income groups
weather_income <- weather |>
  filter(!is.na(`Household income`)) |>
  group_by(`Household income`) |>
  summarize(`Checks Daily Weather` = mean(as.logical(`Checks daily weather`) * 100)) |>
  mutate("Doesn't Check Daily Weather" = 100 - `Checks Daily Weather`)

weather_income |>
  kbl(caption = "Percentage of Users Across Household Income Groups who Check Daily Weather") |>
  kable_material(c("striped", "hover"))
```

Table 6: Percentage of Users Across Household Income Groups who Check Daily Weather

Household income	Checks Daily Weather	Doesn't Check Daily Weather
$0 to $9,999	73.33333	26.66667
$10,000 to $24,999	79.01235	20.98765
$25,000 to $49,999	81.06061	18.93939
$50,000 to $74,999	81.98198	18.01802
$75,000 to $99,999	84.61538	15.38462
$100,000 to $124,999	79.80769	20.19231
$125,000 to $149,999	71.42857	28.57143
$150,000 to $174,999	84.61538	15.38462
$175,000 to $199,999	82.60870	17.39130
$200,000 and up	89.65517	10.34483
Prefer not to answer	78.10651	21.89349

```
# Plot percentage of users who check daily weather by household income
ggplot(weather_income, aes(x = `Checks Daily Weather`, y = fct_rev(`Household income`))) +
  geom_bar(stat = 'identity', fill = "seagreen") +
  scale_x_continuous(breaks = seq(0, 100, by = 10),
                     labels = seq(0, 100, by = 10),
                     limits = c(0, 100)) +
  labs(
    title = "Over 70% check the daily weather report across all household income groups",
    subtitle = "Percentage of respondents who check the daily weather report by household income",
    y = "",
    x = "% of Respondents that Check Daily Weather"
  ) +
  theme_minimal() +
  theme(
    plot.title.position = "plot"
  )
```



Next, we wanted to quantify the relationship between how likely a respondent is to typically check the daily weather report based on their household income. We created a logistic regression using the household income groups to predict the likelihood a person is to typically check the daily weather report.

```
# log regression model for household income

#remove prefer not to answers from model
weather_income <- weather |>
  filter(`Household income` != "Prefer not to answer")

lr_income <- logistic_reg() |>
  set_engine("glm") |>
  fit(`Checks daily weather` ~ `Household income`, data = weather_income, family = "binomial")

lr_income |>
  tidy() |>
  select(term, estimate) |>
  kbl(caption = "Coefficient Estimates of Logistic Regression by Household Income") |>
  kable_material(c("striped", "hover"))
```

We get the following regression equation from our logistic regression model.

$\log(p/(1-p)) = 1.012 + 0(\$0 - \$9{,}999) + 0.314(\$10{,}000 - \$24{,}999) + 0.442(\$25{,}000 - \$49{,}999) + 0.504(\$50{,}000 - \$74{,}999) + 0.693(\$75{,}000 - \$99{,}999) + 0.363(\$100{,}000 - \$124{,}999) - 0.095(\$125{,}000 - \$149{,}999) + 0.693(\$150{,}000 - \$174{,}999) + 0.547(\$175{,}000 - \$199{,}999) + 1.148(\$200{,}000 \text{ and up})$

According to the model, the log odds of the respondent checking the weather has an intercept of 1.012, and decreases by 0.095 units if the respondent has an income within \$125,000 - \$149,999. It increases by

Table 7: Coefficient Estimates of Logistic Regression by Household Income

term	estimate
(Intercept)	1.0116009
‘Household income‘$10,000 to $24,999	0.3140688
‘Household income‘$25,000 to $49,999	0.4423521
‘Household income‘$50,000 to $74,999	0.5035263
‘Household income‘$75,000 to $99,999	0.6931472
‘Household income‘$100,000 to $124,999	0.3627173
‘Household income‘$125,000 to $149,999	-0.0953102
‘Household income‘$150,000 to $174,999	0.6931472
‘Household income‘$175,000 to $199,999	0.5465437
‘Household income‘$200,000 and up	1.1478833

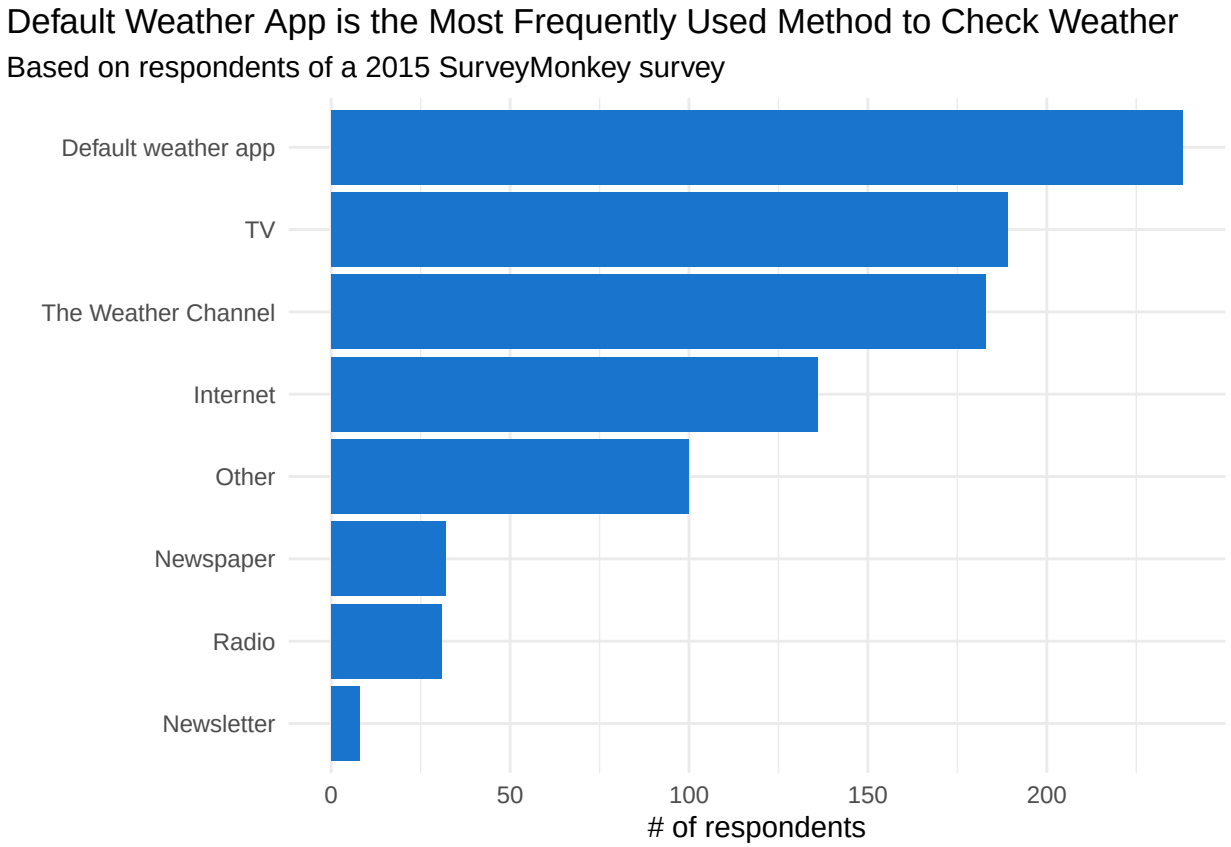
0.314 units, 0.442 units, 0.504 units, 0.693 units, 0.363 units, 0.693 units, 0.547 units, and 1.148 units if the respondent has an income within $10,000 - $24,999, $25,000 - $49,999, $50,000 - $74,999, $75,000 - $99,999, $100,000 - $124,999, $150,000 - $174,999, $175,000 - $199,999, and $200,000 and up respectively. The log odds were unaffected by the respondent having an income less than $9,999 since that was the base level for our model. It seems that users are the most likely to check a daily weather report if they are in the $200,000 and up group, whereas being in the $125,000 to $149,999 group are least likely to check a daily weather report.

Next, we wanted to answer

> What is the most popular website/app used to check the weather?

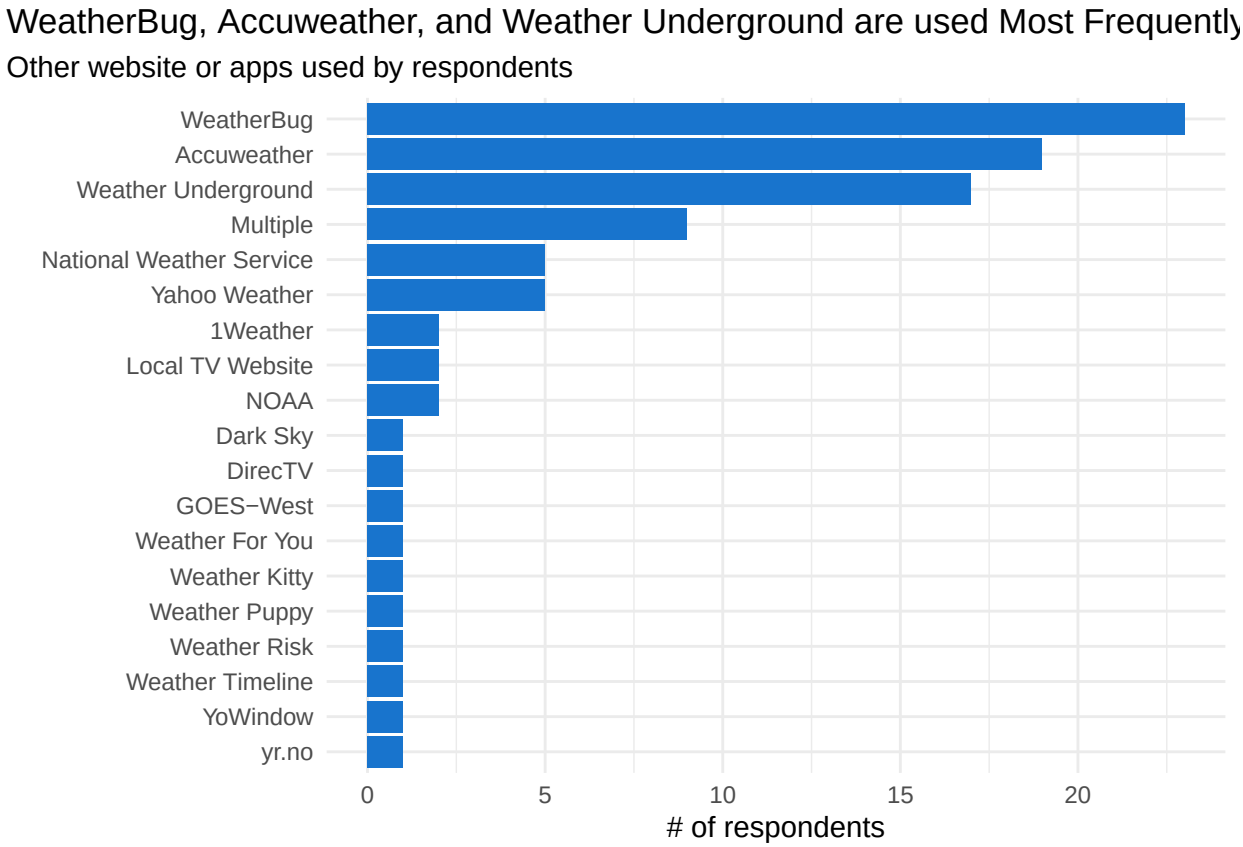
First, we plotted the counts of each method that users could select from. Based on this bar graph, the default weather app was used most commonly to check weather. The Weather Channel and TV were the next popular methods out of all the respondents of the survey.

```
weather |>
  filter(!is.na(`Method of checking weather`)) |>
  ggplot(aes(fct_rev(fct_infreq(`Method of checking weather`)))) +
    geom_bar(fill = "dodgerblue3") +
    coord_flip() +
    labs(title = "Default Weather App is the Most Frequently Used Method to Check Weather",
         subtitle = "Based on respondents of a 2015 SurveyMonkey survey",
         x = "",
         y = "# of respondents") +
  theme_minimal() +
  theme(plot.title.position = "plot")
```



Additionally, we looked at the other website or apps the respondents also used that weren't included in the original choices. We found that WeatherBug, Accuweather, and Weather Underground were the top three other websites and apps used. There were also a number of respondents who said that they used multiple different sources for the daily weather report, as well as many unique website or apps that weren't as common.

```
weather |>
  filter(!is.na(`Other website or app`)) |>
  ggplot(aes(fct_rev(fct_infreq(`Other website or app`)))) +
    geom_bar(fill = "dodgerblue3") +
    coord_flip() +
    labs(title = "WeatherBug, Accuweather, and Weather Underground are used Most Frequently",
         subtitle = "Other website or apps used by respondents",
         x = "",
         y = "# of respondents") +
  theme_minimal() +
  theme(plot.title.position = "plot")
```



Lastly, we also wanted to investigate the following question.

> Can we predict whether a person typically checks the daily weather report based on the method they use to check the weather and their demographic information?

Using most of the features in our data set, including the method respondents chose to use to check the weather, whether respondents would check the weather on a smart watch, and their demographic information, we sought to determine if we could predict whether the respondent typically checks a daily weather report. We found that the best model, chosen by using the least AIC value, encompasses what method the respondent uses to check the weather, whether the respondent would check the weather on their smartwatch, their age, and the region they live in.

```
# run log regression model
weather_reg <- logistic_reg() |>
  set_engine("glm") |>
  fit(`Checks daily weather` ~ `Method of checking weather`
      + `Check weather on smartwatch`
      + Age
      + Region,
      data = weather, family = "binomial")
```

```
# Display table of coefficients
weather_reg |>
  tidy() |>
  select(term, estimate) |>
  kbl(caption = "Coefficient Estimates of Best Logistic Regression Model") |>
  kable_material(c("striped", "hover"))
```

Table 8: Coefficient Estimates of Best Logistic Regression Model

term	estimate
(Intercept)	2.0123613
‘Method of checking weather‘Internet	-0.3683466
‘Method of checking weather‘Newsletter	14.5055044
‘Method of checking weather‘Newspaper	-0.1648286
‘Method of checking weather‘Other	0.7020521
‘Method of checking weather‘Radio	-0.3591698
‘Method of checking weather‘The Weather Channel	0.4906256
‘Method of checking weather‘TV	-0.5242086
‘Check weather on smartwatch‘Somewhat likely	-1.3951164
‘Check weather on smartwatch‘Somewhat unlikely	-2.0118743
‘Check weather on smartwatch‘Very unlikely	-1.3850197
Age30 - 44	0.6091256
Age45 - 59	0.9981112
Age60+	1.4456172
RegionEast South Central	0.9898725
RegionMiddle Atlantic	0.2541885
RegionMountain	-0.4128652
RegionNew England	1.2830533
RegionPacific	-0.9056831
RegionSouth Atlantic	-0.7991243
RegionWest North Central	-0.0403200
RegionWest South Central	0.7321553

```
# AIC value
weather_reg |>
  glance() |>
  pull(AIC)
```

```
## [1] 765.4523
```

After finding the best logistic regression model based on AIC value, we get the following model:

(Each line of the following model represents one variable; in order, it is `Method of checking weather`, `Check weather on smartwatch`, Age, Region)

log(p/(1-p)) =

2.012 +

0(Default Weather App) + 14.506(Newsletter) + 0.702(Other) + 0.491(Weather Channel) - 0.368(Internet) - 0.165(Newspaper) - 0.359(Radio) - 0.524(TV) +

0(Very Likely) - 1.395(Somewhat Likely) - 2.012(Somewhat Unlikely) - 1.385(Very Likely) +

0(Age 18-29) + 0.609(Age 30-44) + 0.998(Age 45-59) + 1.446(Age 60+) +

0(East North Central) + 0.990(East South Central) + 0.254(Middle Atlantic) - 0.413(Mountain) + 1.283(New England) - 0.906(Pacific) - 0.799(South Atlantic) - 0.040(West North Central) + 0.732 (West South Central)

According to the best-fit model, the log odds of a person from this survey checking the weather were the highest if they were using a newsletter, were likely to check the weather on their smartwatch, were over the age of 60, and lived in New England. On the flip side, a person from this survey was least likely to check the weather if they used a radio, were somewhat unlikely to check the weather on their smartwatch, were

between the ages of 18-29, and lived in the Pacific region. Their gender and household income seemed to have little effect on whether they checked the weather or not. However, because this survey was not weighted, we cannot be sure of how representative this model is.

Results

We began our analysis by exploring the data collected from the survey. Through various plots, we visualized the distribution of our data to get a better understanding of our data for our analysis. We found that there were a significant number of people who indicated that they do typically check a daily weather report compared to those who didn't. There were also a good range of respondents' ages, ranging from 18 to over 60. Most of the respondents in our dataset were over 44 years old. We also saw that most of the respondents were female, with over a difference of 100 respondents compared to males. We also explored the distribution of reported household income ranges and found a majority of respondents fell between the range of $10,000 to $124,999, with a considerable number of outliers in the $200,000 and up group. Lastly, we also looked into the distribution of respondents' regions and found that most respondents are located near the East and West Coast, with the most respondents coming from the Pacific Region.

Our first analysis looked into the relationship between sex and whether a person typically checks a daily weather report. We plotted the percentages of males and females who indicated that they checked a daily weather report and found that the two groups were fairly similar, with Males having a slightly higher percentage of respondents indicating that they check a daily weather report. Additionally, we explored the difference between sex and how they checked a daily weather report. Across all methods recorded, both sexes had a similar relative distribution, with the default weather app, TV, and The Weather Channel being among the top three most popular methods for both groups. Lastly, we ran a logistic regression model using sex to predict whether the respondent typically checks the daily weather report. We found that males had a higher likelihood of checking the weather report, but with only a small increase of 0.131 in the log likelihood value compared to female. From our results, we can conclude that there is no strong relationship between whether one checks the weather report and the sex, however, males were only slightly more likely to do so compared to females.

The next analysis explored whether weather app use differed based on the respondent's region. We plotted the percentages of those who marked that they did check daily weather reports versus those who didn't within each region. New England had the highest percentage, while most of the Central regions had the next highest percentages. Interestingly, some of the coastal regions had much lower percentages compared to the other regions, although most regions had over 80% respondents indicating that they do check daily weather reports. We also ran a logistic regression model to quantify the likelihood of checking daily weather reports based on region and found that New England was most likely, while all other regions were less likely, with the least likely region being the Pacific region. Furthermore, we also explored the differences between regions and methods of checking the weather report. While all the regions did differ in their most popular method of checking the weather, the default weather app, The Weather Channel, TV, and the Internet were the most popular methods across most regions. Weather app usage did differ based on geographic regions. However, nearby geographical regions didn't always share a similar trend.

Next, we analyzed whether there was a relationship between household income and checking the weather app. We found that all recorded household income groups had over 70% of respondents indicating that they checked a daily weather report. We also applied a logistic regression model to quantify the likelihood that someone checks a daily weather report based on their household income. From our analysis, we determined that respondents from households that make $200,000 and up were the most likely to check a daily weather report, while those from the $125,000 to $149,999 group were the least likely group to check a daily weather report.

We then analyzed what the most popular website/app used was to check the weather. Through plots of the counts of each category and responses from the other website or apps that respondents could include other options, we identified the most frequently used sources to check for weather reports. From the original response options in the survey, the default weather app, TV, and The Weather Channel were the most popular methods of checking the weather report. Additionally, within the other options that users could fill

out, WeatherBug, Accuweather, and Weather Underground were used the most frequently.

Lastly, we also wanted to determine if we could predict how likely a person is to typically check a daily weather report. Initially, we sought to use all features from our dataset to make the prediction through a logistic regression. However, we performed backwards elimination to identify the best subset of features that yielded the best model based on AIC value. The model we identified best predicted whether a person was likely to typically check a daily weather report used the method the respondent used to check the weather, whether the respondent would check the weather on a smartwatch, their age, and the region they live in. We also identified situations that yielded the highest and lowest likelihood of checking a daily weather report based on the coefficients of our model.

Conclusion

This analysis was originally inspired by this Wall Street Journal piece and a fivethirtyeight article discussing Verizon Fios's decision to drop The Weather Channel from their channel offerings. With the increasing impact of climate change and its effects on weather patterns, we believe that it is more important than ever for individuals to stay informed about current weather conditions. To extend their analyses, we wanted to use their data to identify patterns in users' weather checking behaviors. We found that the majority of Americans check the weather at least once a day, with the most common method being through a mobile app. The study also highlighted regional differences in weather checking habits, with residents of the Midwest, East, and South being more likely to check the weather multiple times a day.

However, our analysis may be limited, as the data was collected in 2015. It is also notable that younger generations rely more on social media for weather updates. Additionally, our study did not account for extreme weather events, such as hurricanes or severe winter storms, which can greatly increase the frequency of weather checks. An extension of our analysis would likely involve creating a new survey, so we can record changes in American's weather-checking behavior.

Overall, this report emphasizes the importance of staying informed about weather conditions and understanding regional differences in weather checking habits. With the growing threat of climate change and its impacts on our weather, it is crucial for individuals to take a proactive approach in monitoring weather patterns and preparing for potential risks.