

Geographical Maps with ggplot2

```
# load dplyr package
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

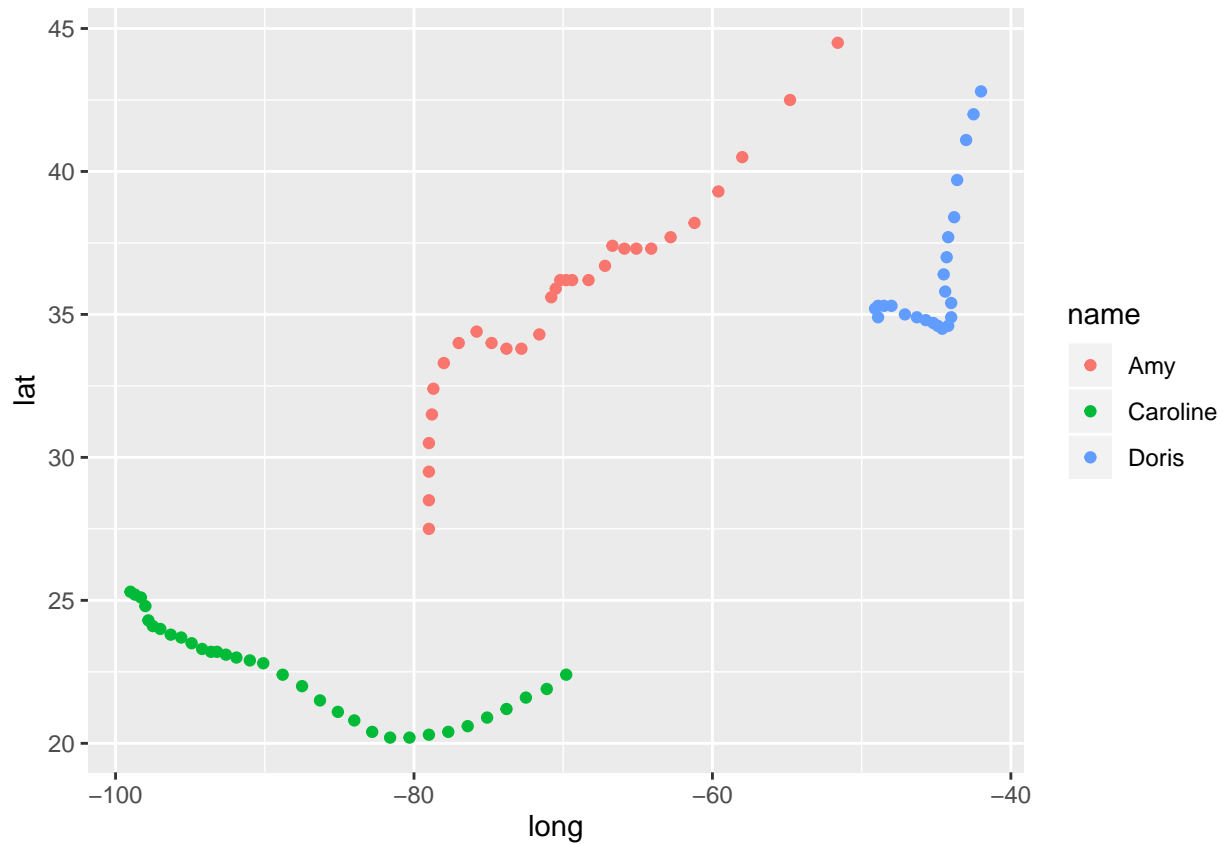
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
storms75 <- filter(storms, year == 1975)
head(storms75)
```

```
## # A tibble: 6 x 13
##   name   year month   day hour   lat   long status category  wind pressure
##   <chr> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <chr>   <ord>    <int>    <int>
## 1 Amy    1975     6    27     0  27.5 -79   tropi~ -1      25     1013
## 2 Amy    1975     6    27     6  28.5 -79   tropi~ -1      25     1013
## 3 Amy    1975     6    27    12  29.5 -79   tropi~ -1      25     1013
## 4 Amy    1975     6    27    18  30.5 -79   tropi~ -1      25     1013
## 5 Amy    1975     6    28     0  31.5 -78.8 tropi~ -1      25     1012
## 6 Amy    1975     6    28     6  32.4 -78.7 tropi~ -1      25     1012
## # ... with 2 more variables: ts_diameter <dbl>, hu_diameter <dbl>
```

```
ggplot(storms75, aes(x = long, y = lat)) + geom_point(aes(color = name))
```



```
# alternative ways to write equivalent commands
# ggplot(data = storms75) + geom_point(aes(x = long, y = lat, color = name))
# ggplot() + geom_point(data = storms75, aes(x = long, y = lat, color = name))
```

But where is it? We need an image of world map.

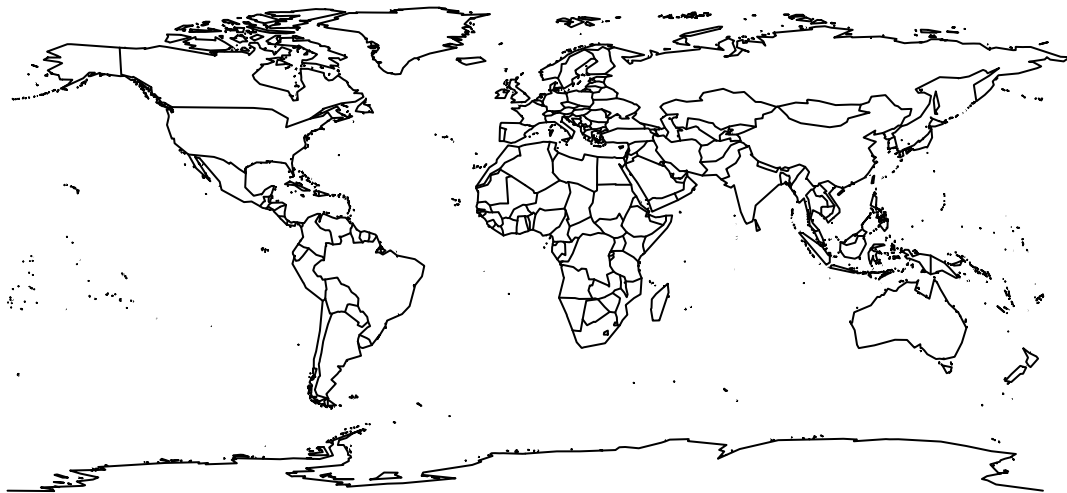
Using maps package

```
# install maps package
# install.packages("maps")

# load maps package
library(maps)
```

map() plots geographical maps.

```
# world map
map("world")
```



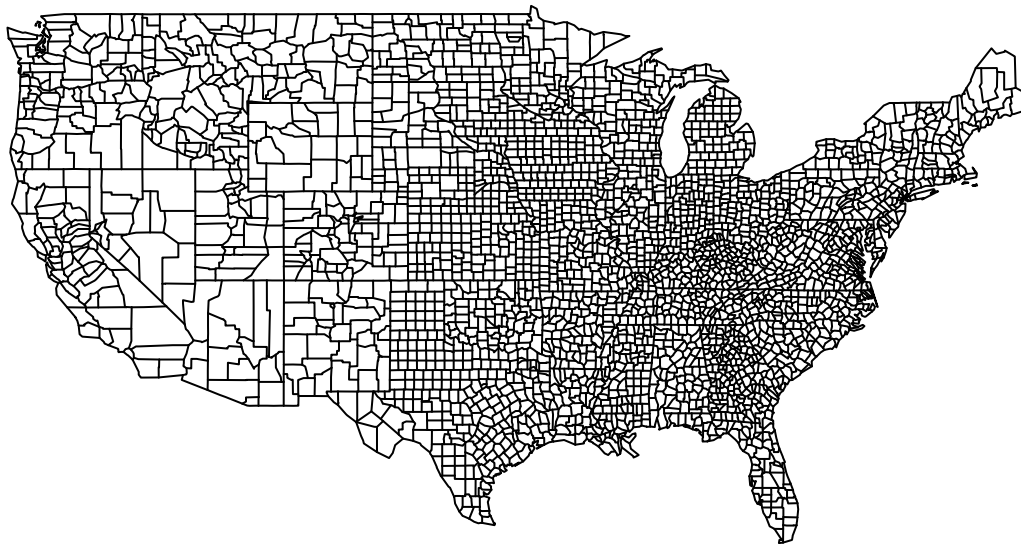
```
# usa map  
map("usa")
```



```
# usa states  
map("state")
```



```
# usa counties
map("county")
```



To use `ggplot2` to map map objects, first we need to convert a map object to a dataframe object.

```
world_map <- map_data("world")
head(world_map)
```

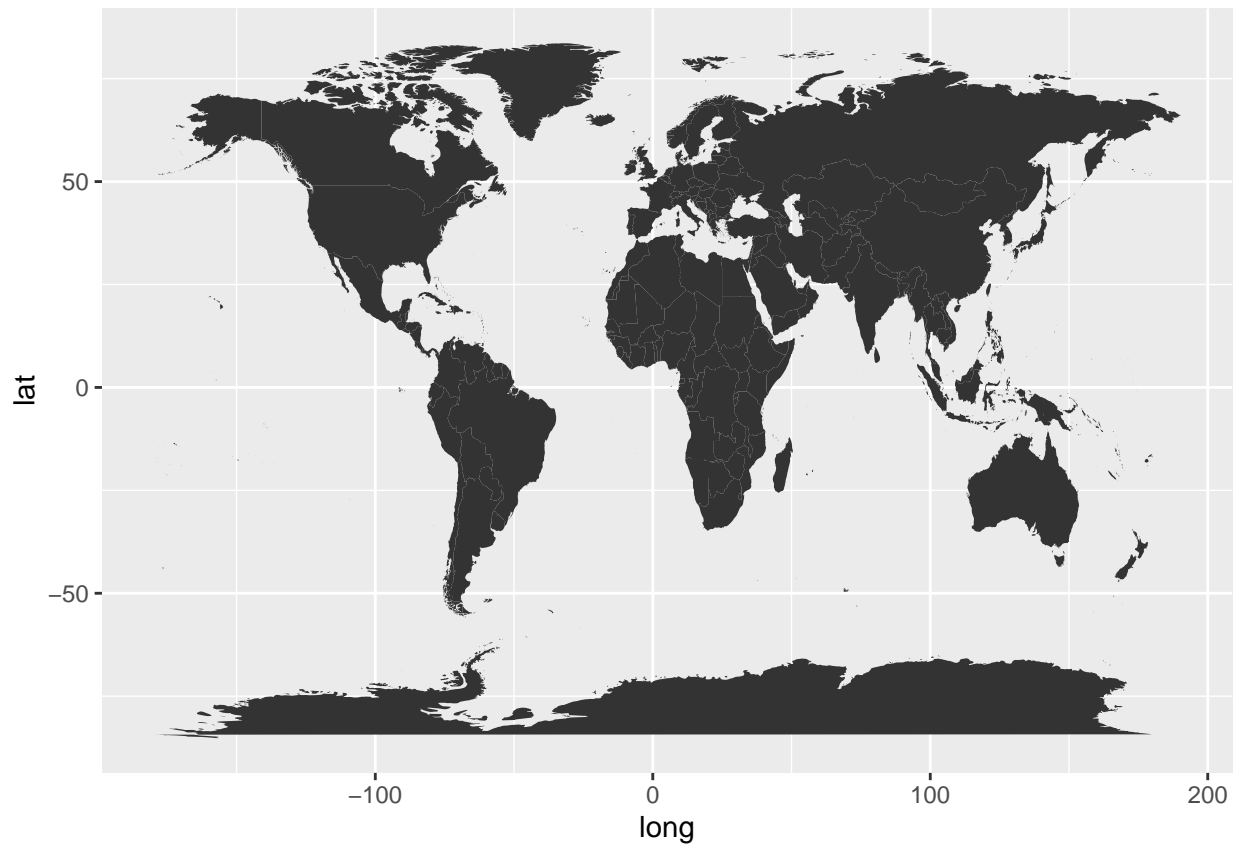
```
##      long      lat group order region subregion
## 1 -69.89912 12.45200     1     1  Aruba      <NA>
## 2 -69.89571 12.42300     1     2  Aruba      <NA>
## 3 -69.94219 12.43853     1     3  Aruba      <NA>
## 4 -70.00415 12.50049     1     4  Aruba      <NA>
## 5 -70.06612 12.54697     1     5  Aruba      <NA>
## 6 -70.05088 12.59707     1     6  Aruba      <NA>
```

```
# check if world_map is really a dataframe object  
class(world_map)
```

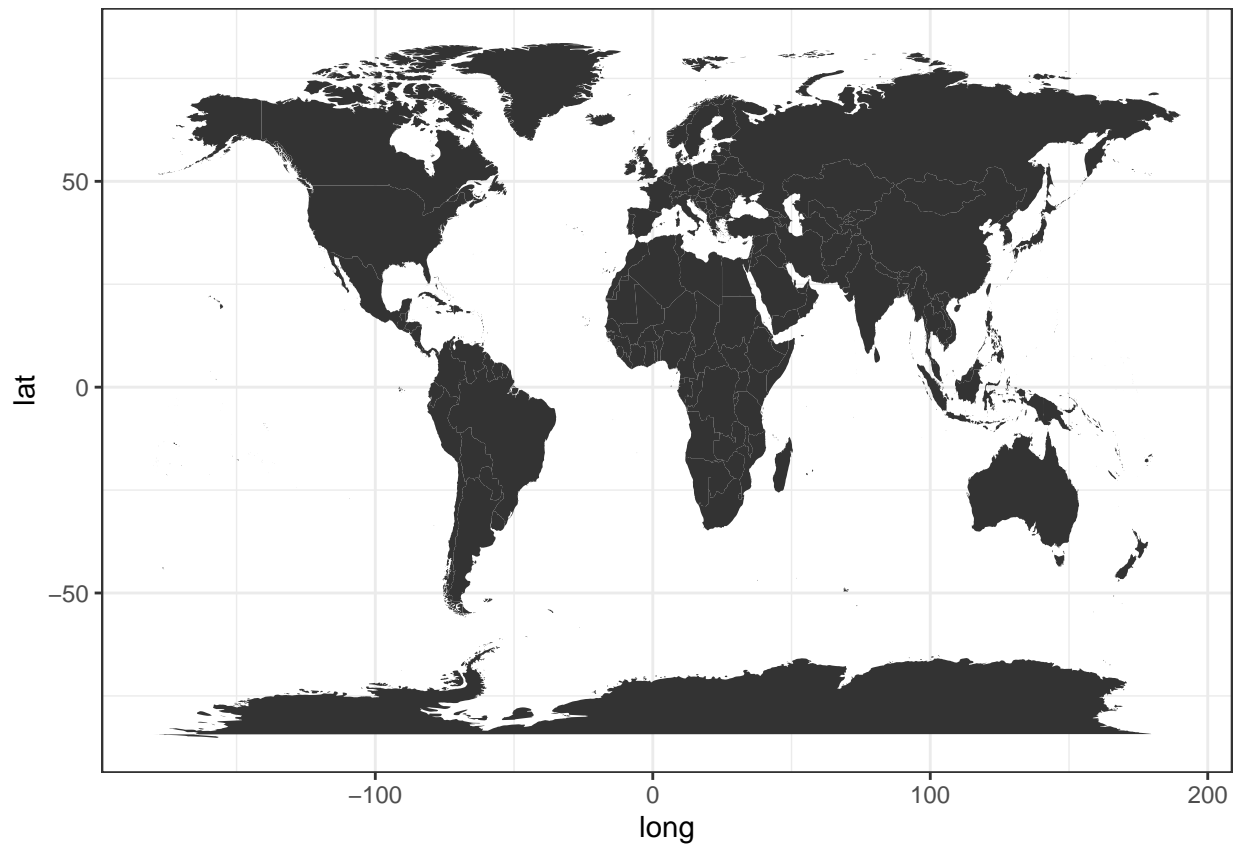
```
## [1] "data.frame"
```

Use `geom_polygon()` to draw a map. `theme_bw()` changes the map to black-white.

```
ggplot() +  
  geom_polygon(data = world_map, aes(x = long, y = lat, group = group))
```

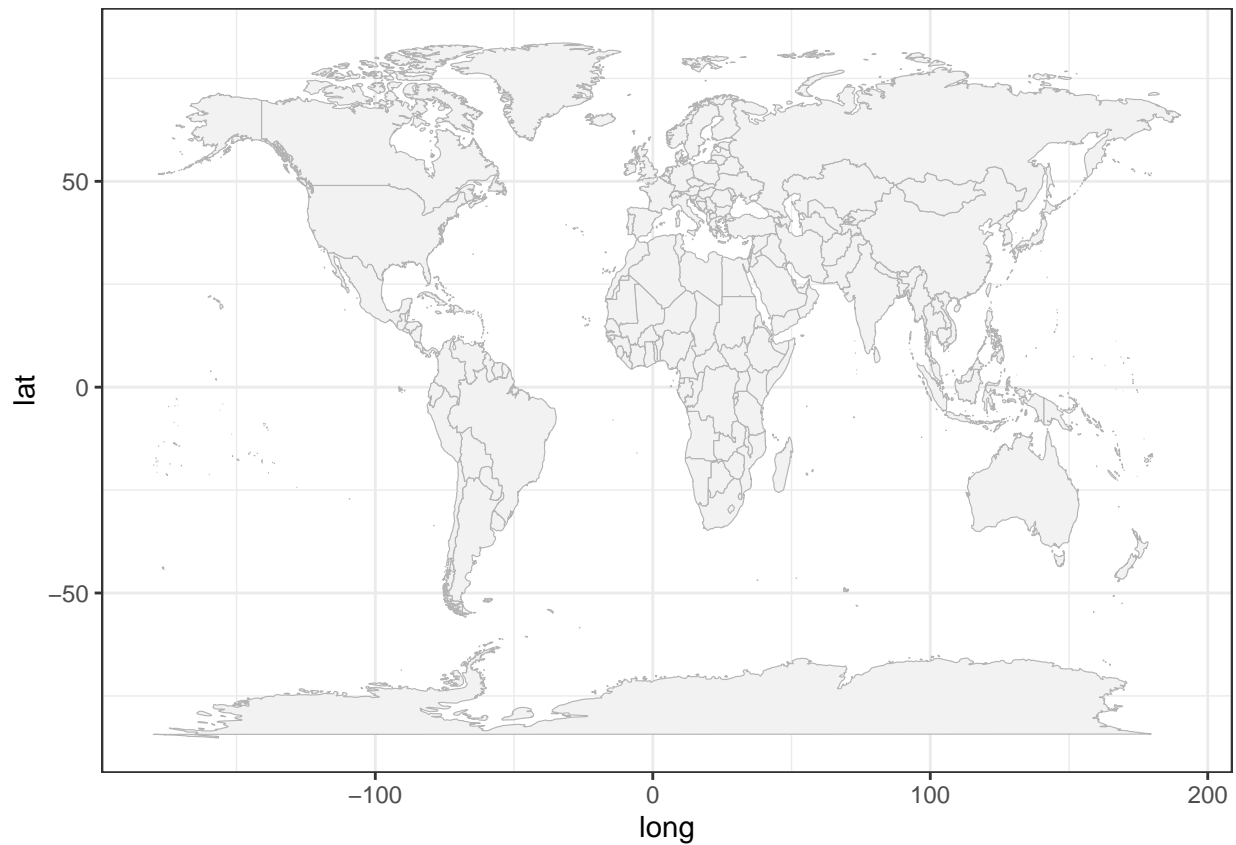


```
# use theme_bw()  
# the background changes to white  
ggplot() +  
  geom_polygon(data = world_map, aes(x = long, y = lat, group = group)) +  
  theme_bw()
```



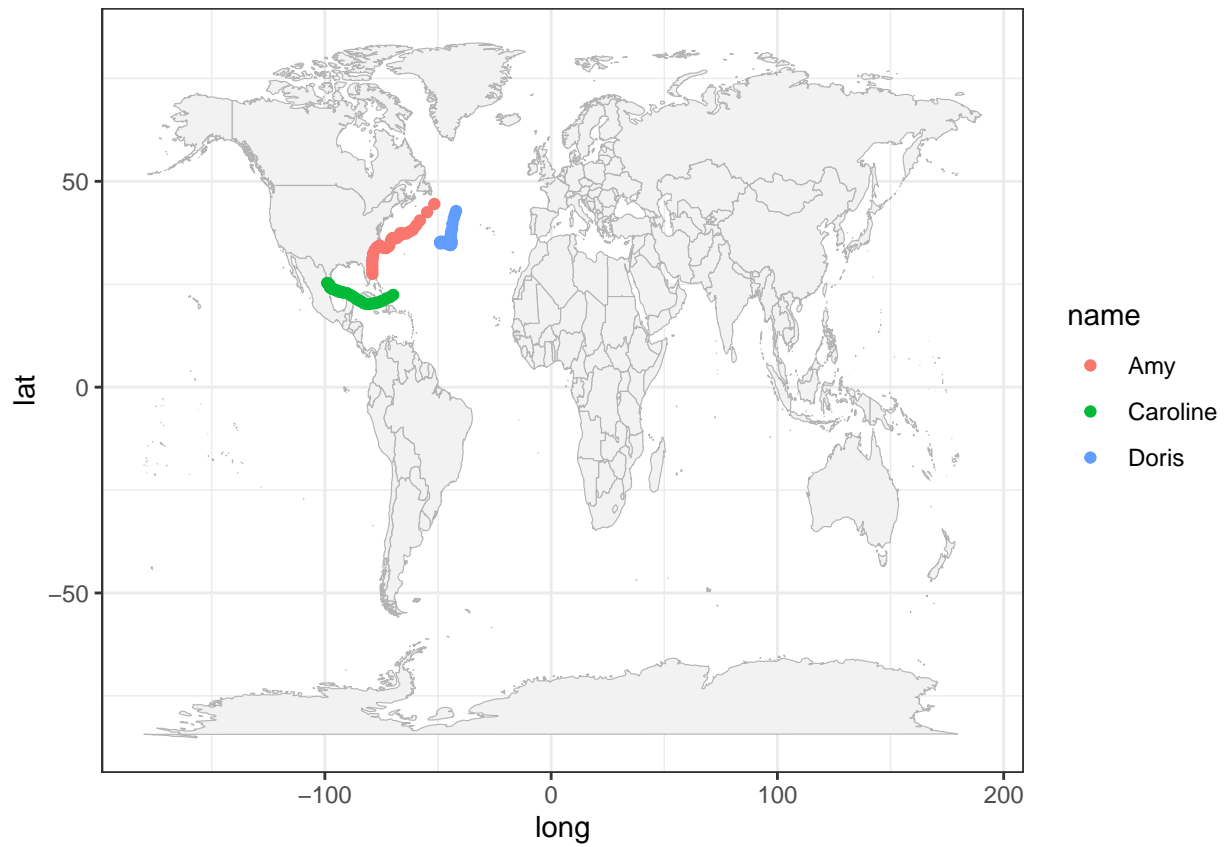
```
# store the ggplot to a variable
gg_world <- ggplot() +
  geom_polygon(data = world_map,
    aes(x = long, y = lat, group = group),
    fill = "gray95", colour = "gray70", size = 0.2) +
  theme_bw()

gg_world
```

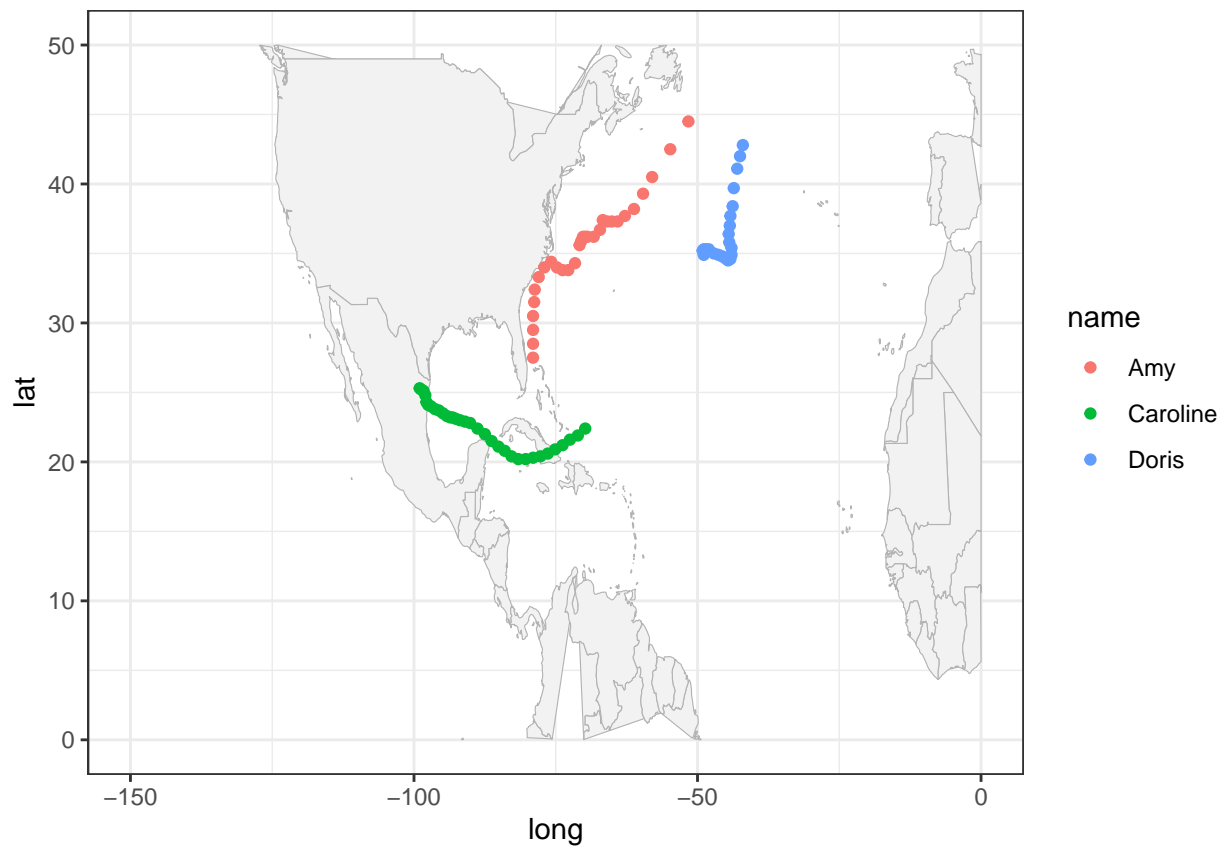


Let's add the points of the storm records. Use `xlim()` and `ylim()` to specify the range of longitude and latitude for zooming-in.

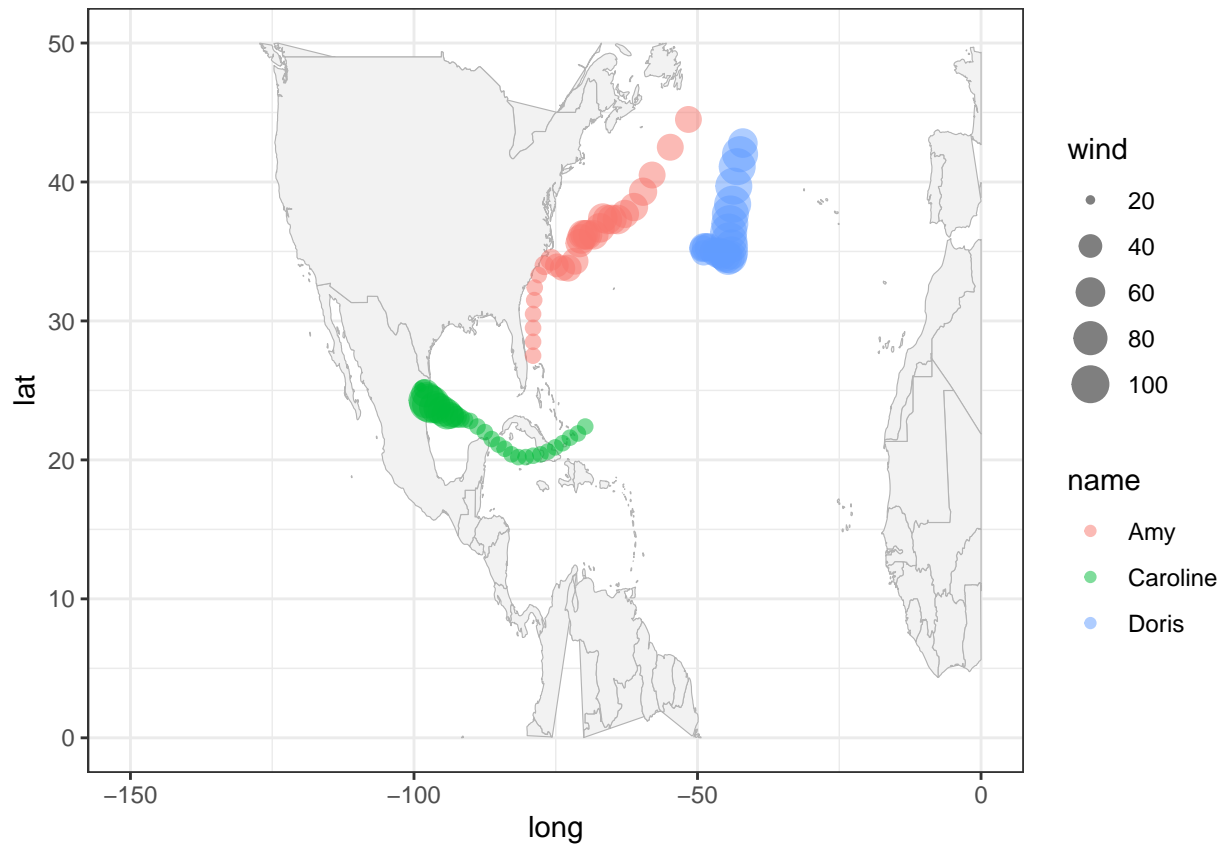
```
gg_world + geom_point(data = storms75, aes(x = long, y = lat, color = name))
```



```
# zoom-in
gg_world + geom_point(data = storms75, aes(x = long, y = lat, color = name)) +
  xlim(c(-150, 0)) +
  ylim(c(0, 50))
```

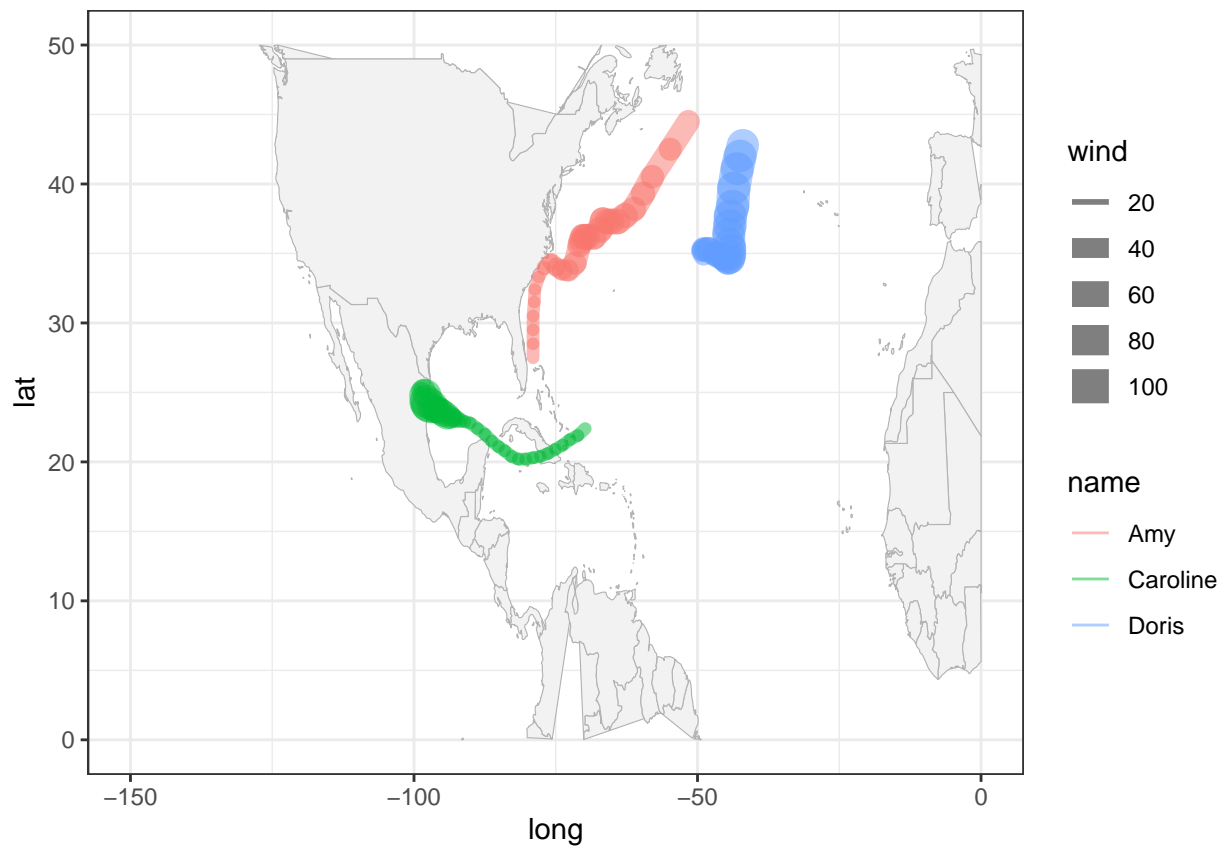



```
# add wind speed feature by setting size argument
gg_world +
  geom_point(data = storms75,
             aes(x = long, y = lat, color = name, size = wind), alpha = 0.5) +
  xlim(c(-150, 0)) +
  ylim(c(0, 50))
```



`geom_path()` acts like `geom_point()`. It connects points but it's more like a "path" of the points.

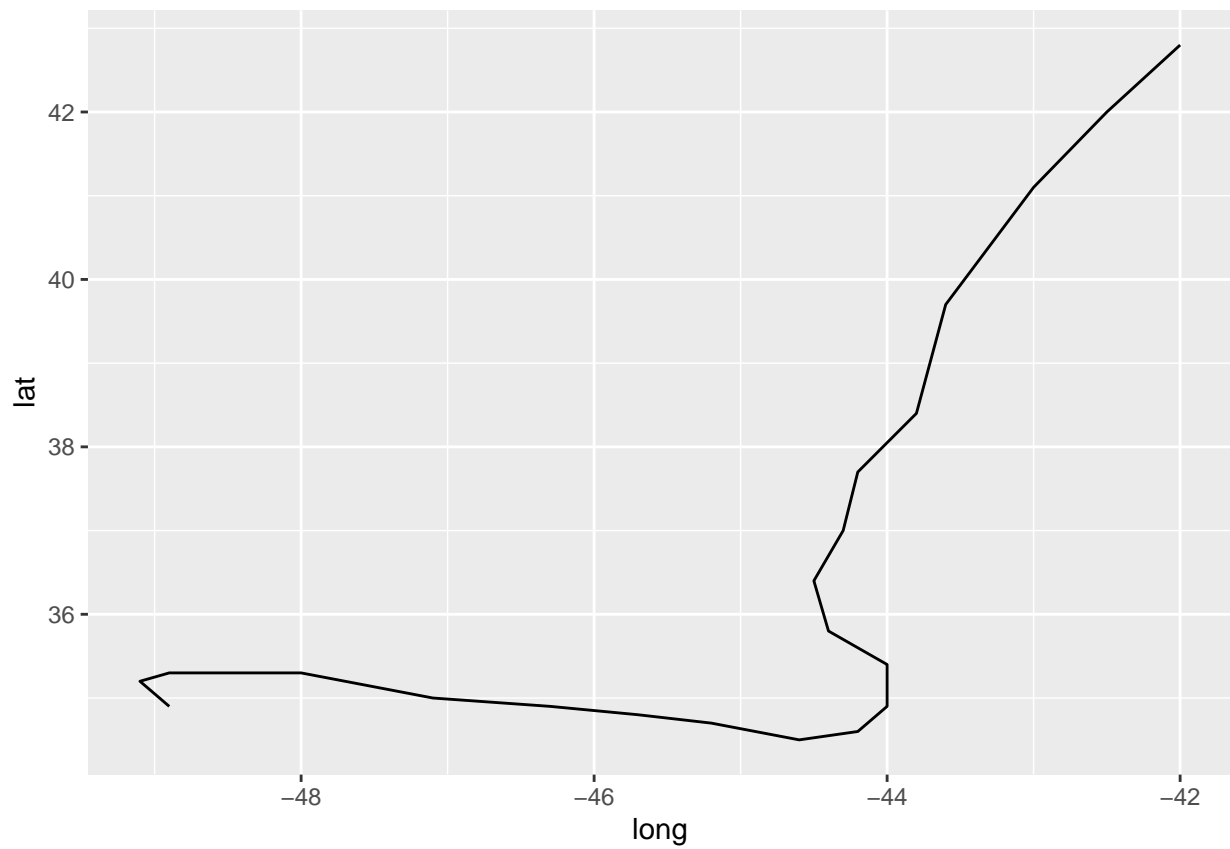
```
# similar result
gg_world +
  geom_path(data = storms75,
            aes(x = long, y = lat, color = name, size = wind),
            alpha = 0.5, lineend = "round") +
  xlim(c(-150, 0)) +
  ylim(c(0, 50))
```



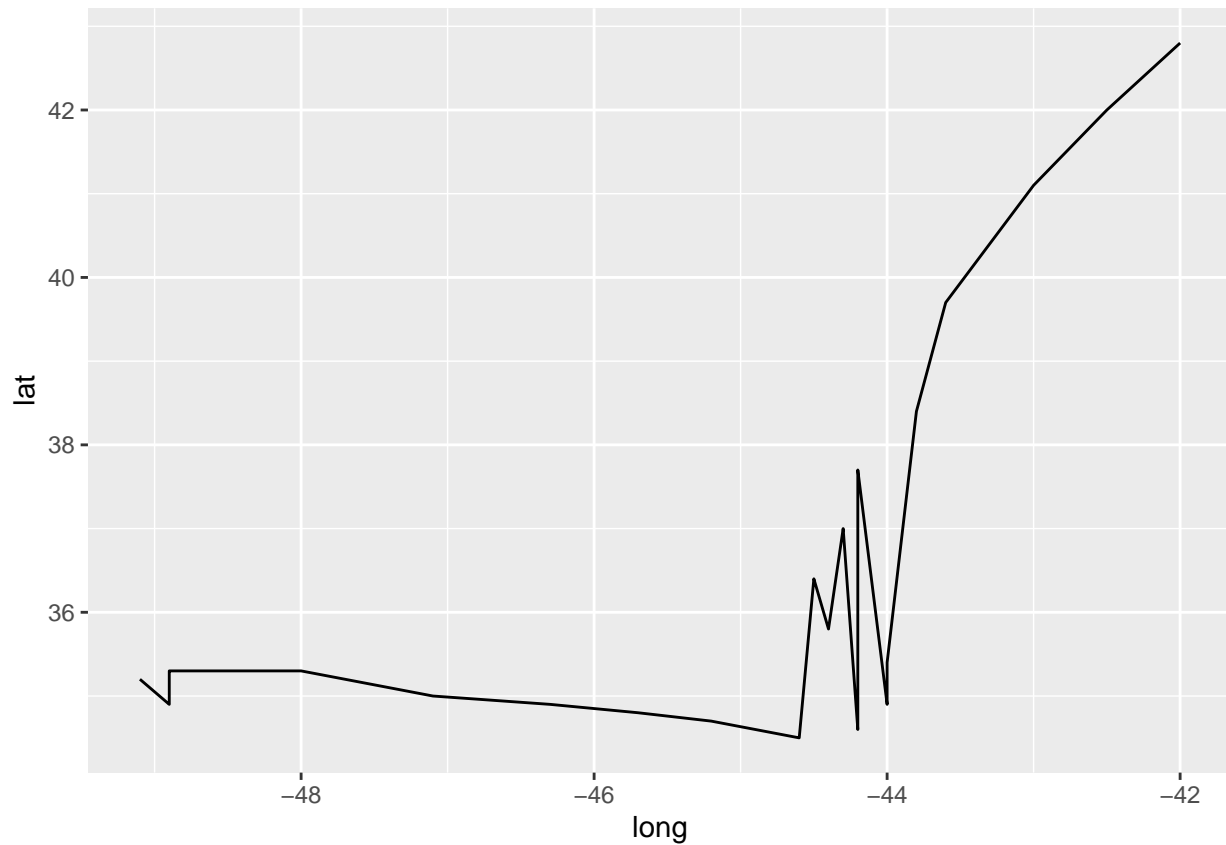
See the difference between `geom_line()` and `geom_path()` Both connects points but there's a big difference.

```
# just an example
doris75 <- filter(storms75, name == "Doris")

# geom_path()
ggplot(data = doris75, aes(x = long, y = lat)) + geom_path()
```



```
# geom_line()
ggplot(data = doris75, aes(x = long, y = lat)) + geom_line()
```



Using `rnaturalearth` package

```
# install rnaturalearth package
# install.packages("rnaturalearth")
# may need additional packages
# install.packages(c("rnaturalearthdata", "rgeos"))

# load rnaturalearth package
library(rnaturalearth)

# library(rnaturalearthdata)
# library(rgeos)
```

`ne_countries()` gets world map polygons.

```
world_df <- ne_countries(scale = "medium", returnclass = "sf")
class(world_df)
```

```
## [1] "sf"          "data.frame"
```

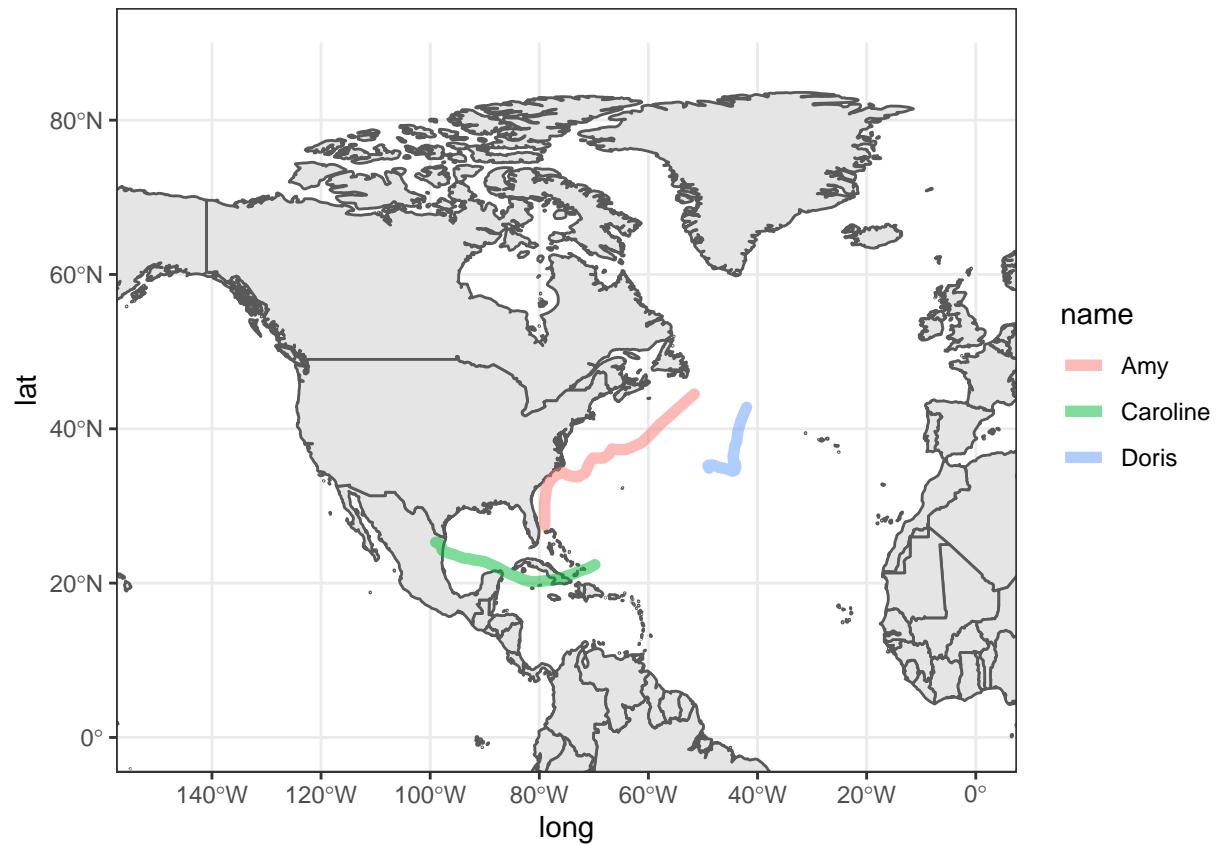
`geom_sf()` visualizes simple feature objects `sf`,

```
ggplot(data = world_df) + geom_sf() + theme_bw()
```



We can set specify the region using `coord_sf()`, instead of using `xlim()` and `ylim()`.

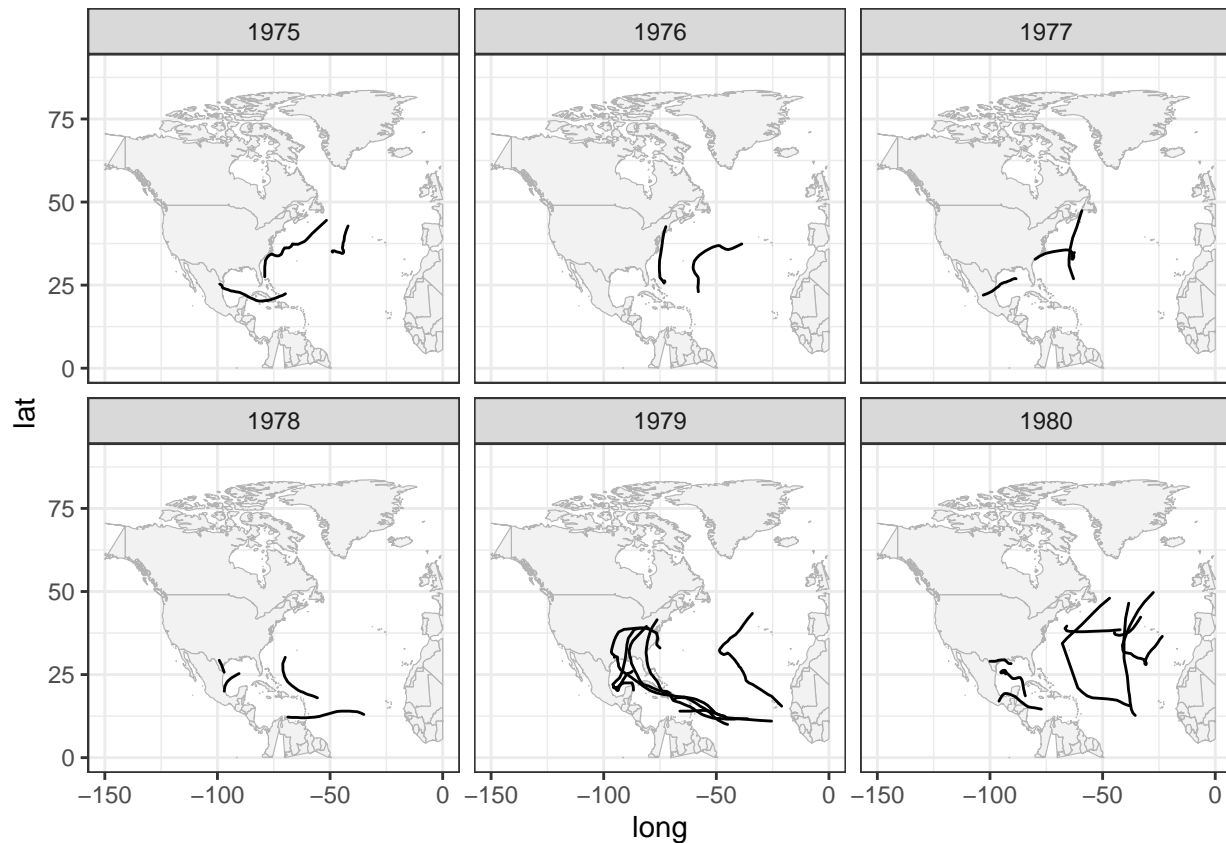
```
gg_world2 <- ggplot(data = world_df) +  
  geom_sf() +  
  coord_sf(xlim = c(-150, 0), y = c(0, 90), expand = TRUE) +  
  theme_bw()  
  
gg_world2 +  
  geom_path(data = storms75,  
    aes(x = long, y = lat, color = name),  
    lineend = "round", size = 2, alpha = 0.5)
```



Facet by year from 1975 to 1980.

```
storms_75_80 <- filter(storms, year %in% 1975:1980)

gg_world + geom_path(data = storms_75_80,
  aes(x = long, y = lat, group = name),
  lineend = "round") +
  xlim(c(-150, 0)) + ylim(c(0, 90)) +
  facet_wrap(~ year)
```



Exercises

- 1) Filter storms in the 1980's decade (1980 - 1989) and make a plot, with facets by month.year

```
storms80s <- filter(storms, year %in% 1980:1989)
head(storms80s)
```

```
## # A tibble: 6 x 13
##   name   year month   day hour   lat long status category  wind pressure
##   <chr> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <chr>   <ord>    <int>    <int>
## 1 Bonn~ 1980     8    14     0  12.7 -35.5 tropi~ -1      25     1010
## 2 Bonn~ 1980     8    14     6  13.5 -36.6 tropi~ -1      30     1008
## 3 Bonn~ 1980     8    14    12  14.7 -37.3 tropi~  0      35     1005
## 4 Bonn~ 1980     8    14    18  15.7 -37.5 tropi~  0      45     1000
## 5 Bonn~ 1980     8    15     0  16.7 -37.8 tropi~  0      45     1000
## 6 Bonn~ 1980     8    15     6  17.7 -37.9 tropi~  0      45     1000
## # ... with 2 more variables: ts_diameter <dbl>, hu_diameter <dbl>
```

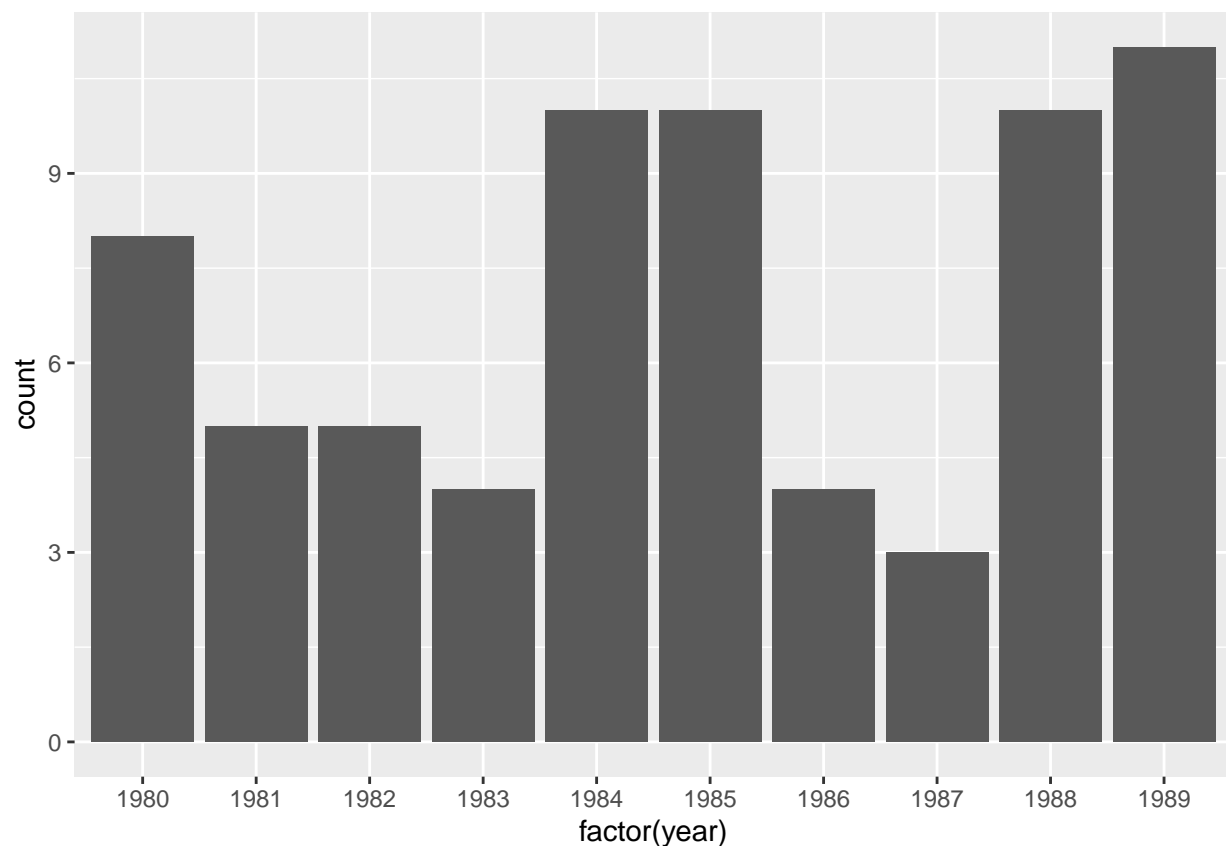
Which year seems to have the largest number of storms? Which year seems to have the smallest number of storms? Does it seem to be a pattern (e.g. increasing number of storms over years)?

```
num_of_storms80s <- distinct(group_by(storms80s, year, name), year)
num_of_storms80s
```



```
## # A tibble: 70 x 2
## # Groups:   year, name [70]
##   year name
##   <dbl> <chr>
## 1 1980 Bonnie
## 2 1980 Charley
## 3 1980 Georges
## 4 1980 Danielle
## 5 1980 Hermine
## 6 1980 Ivan
## 7 1980 Jeanne
## 8 1980 Karl
## 9 1981 Emily
## 10 1981 Floyd
## # ... with 60 more rows
```

```
# by the bar plot, 1989 had the largest number of storms
# and 1987 had the smallest number of storms
ggplot(data = num_of_storms80s, aes(x = factor(year))) + # factor year to set label with only integers
  geom_bar()
```



The number of storms increased since 1980, but there seems no particular pattern over the period.

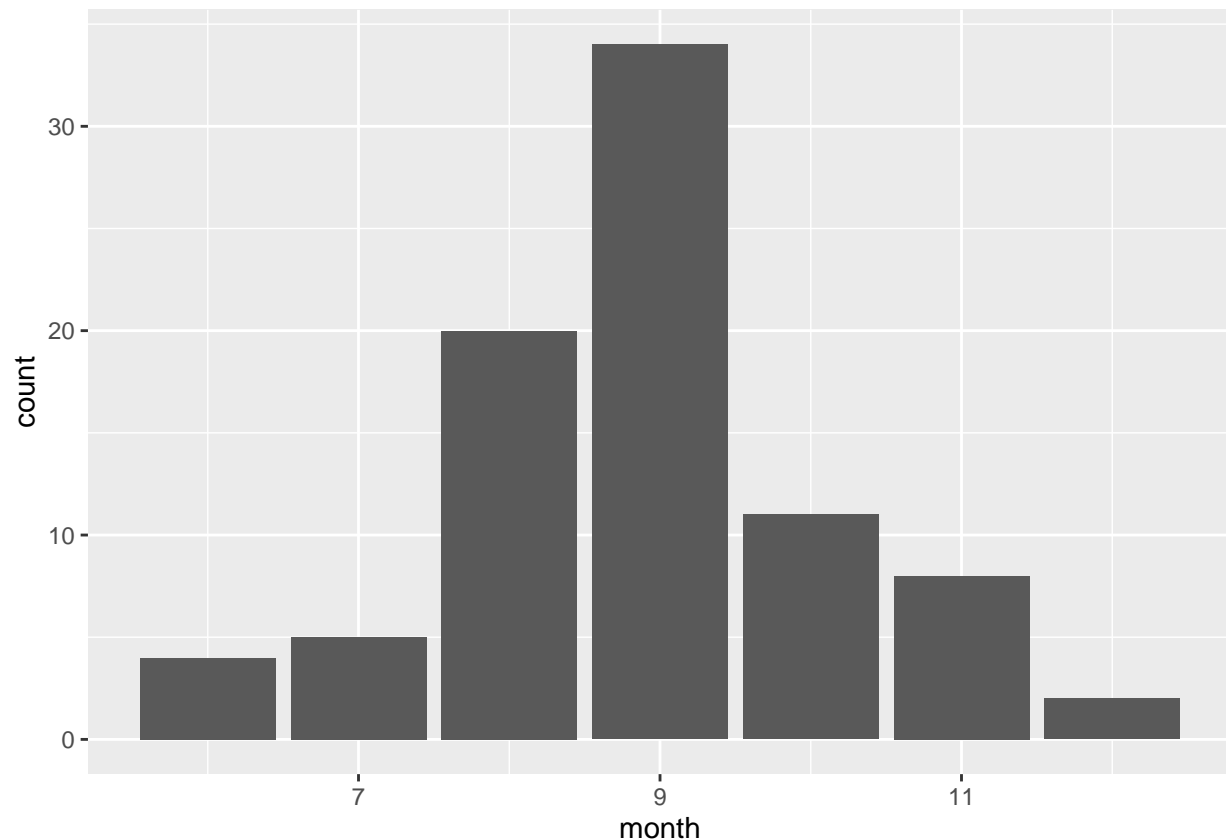
- 2) Take the previous data, storms in the 1980's decade, and make a plot, with but this time with facets by month.

What is the most active month? What is the least active month? Are there months without any storms?

```
num_storms_month <- distinct(group_by(storms80s, year, name, month), month)
num_storms_month
```

```
## # A tibble: 84 x 3
## # Groups:   year, name, month [84]
##   month year name
##   <dbl> <dbl> <chr>
## 1     8  1980 Bonnie
## 2     8  1980 Charley
## 3     9  1980 Georges
## 4     9  1980 Danielle
## 5     9  1980 Hermine
## 6    10  1980 Ivan
## 7    11  1980 Jeanne
## 8    11  1980 Karl
## 9     9  1981 Emily
## 10    9  1981 Floyd
## # ... with 74 more rows
```

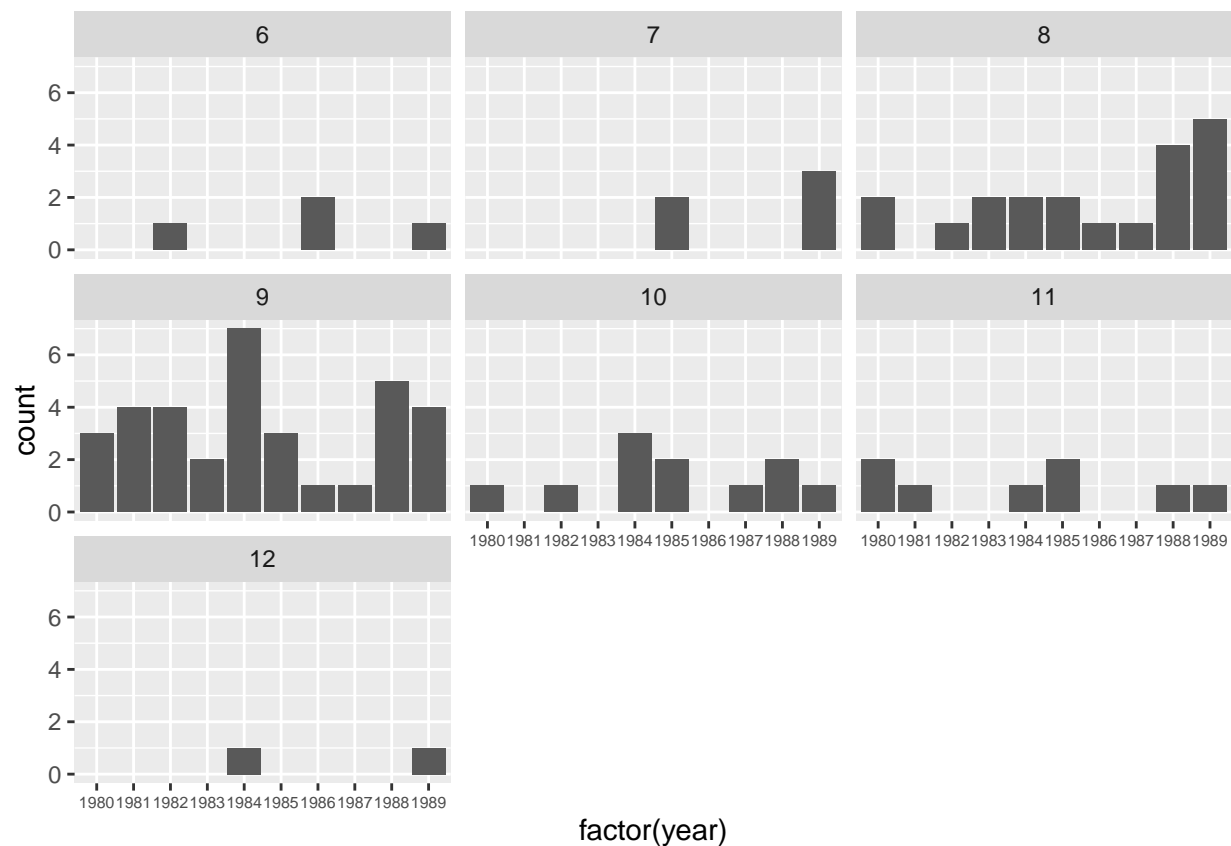
```
ggplot(data = num_storms_month, aes(x = month)) + geom_bar()
```



```
# facet by month but also year
num_storms_year_month <- distinct(group_by(storms80s, year, name, month), year, month)
num_storms_year_month
```

```
## # A tibble: 84 x 3
## # Groups:   year, name, month [84]
##   year month name
##   <dbl> <dbl> <chr>
## 1 1980     8 Bonnie
## 2 1980     8 Charley
## 3 1980     9 Georges
## 4 1980     9 Danielle
## 5 1980     9 Hermine
## 6 1980    10 Ivan
## 7 1980    11 Jeanne
## 8 1980    11 Karl
## 9 1981     9 Emily
## 10 1981     9 Floyd
## # ... with 74 more rows
```

```
ggplot(data = num_storms_year_month, aes(x = factor(year))) +
  geom_bar() +
  facet_wrap(~ month) +
  theme(axis.text.x = element_text(size=6)) # avoid overlapping labels
```



The most active month is September and least active month is December. No storms: January through May.