Case Study in ggplot2

Paul Intrevado

August 02, 2018

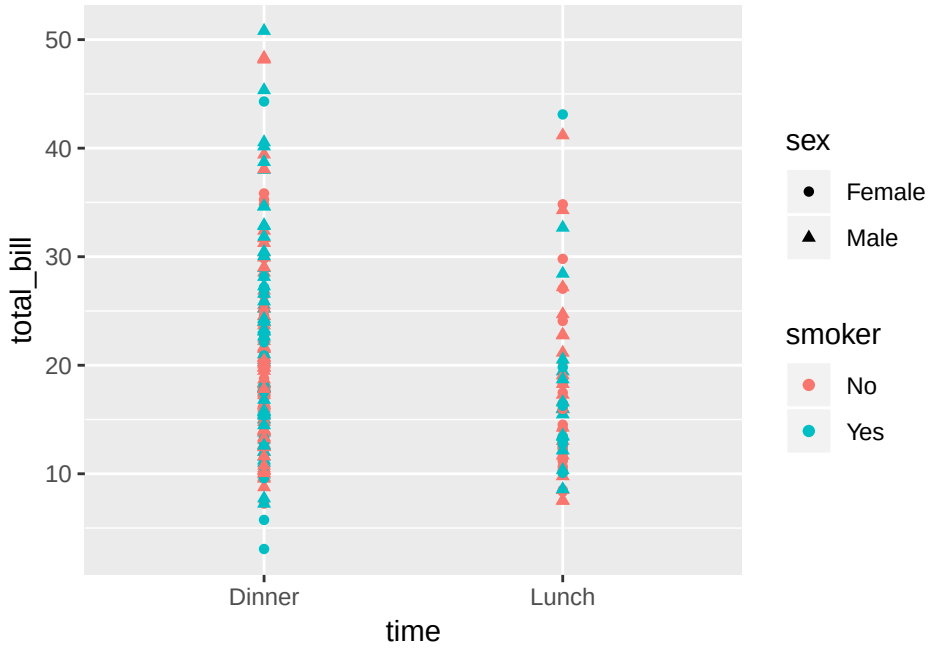
Question 1

The `reshape2` package contains a dataset named `tips`, which contains information on dining trasactions. Summary of the data follows:

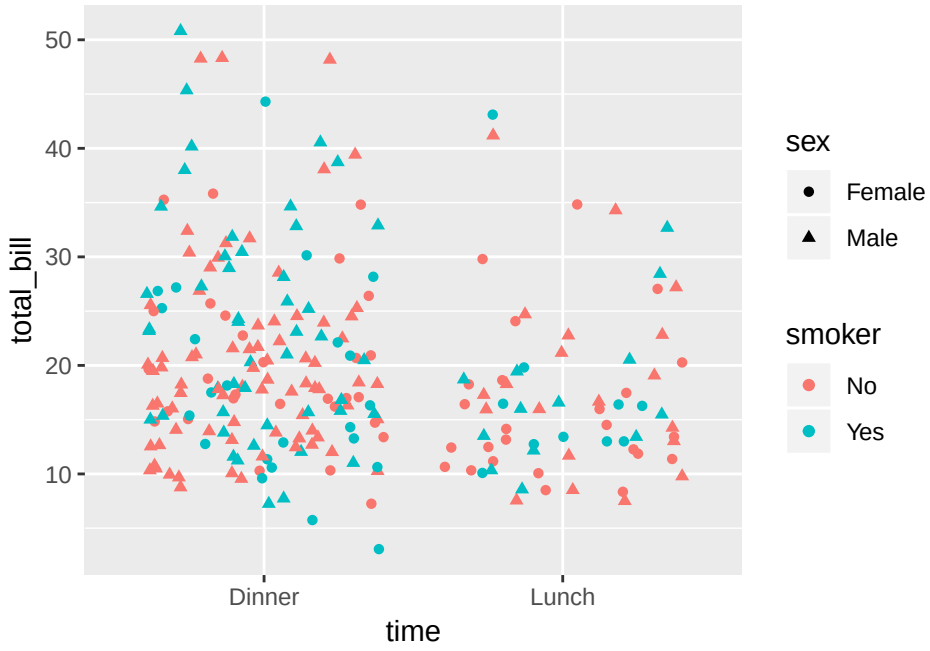
```
## Observations: 244
## Variables: 7
## $ total_bill <dbl> 16.99, 10.34, 21.01, 23.68, 24.59, 25.29, 8.77, 26....
## $ tip        <dbl> 1.01, 1.66, 3.50, 3.31, 3.61, 4.71, 2.00, 3.12, 1.9...
## $ sex        <fct> Female, Male, Male, Male, Female, Male, Male, Male,...
## $ smoker     <fct> No, No, No, No, No, No, No, No, No, No, No, No, No,...
## $ day        <fct> Sun, Sun, Sun, Sun, Sun, Sun, Sun, Sun, Sun, Sun, S...
## $ time       <fct> Dinner, Dinner, Dinner, Dinner, Dinner, Dinner, Din...
## $ size       <int> 2, 3, 3, 2, 4, 4, 2, 4, 2, 2, 2, 4, 2, 4, 2, 2, 3, ...
```

Create the following graphs:

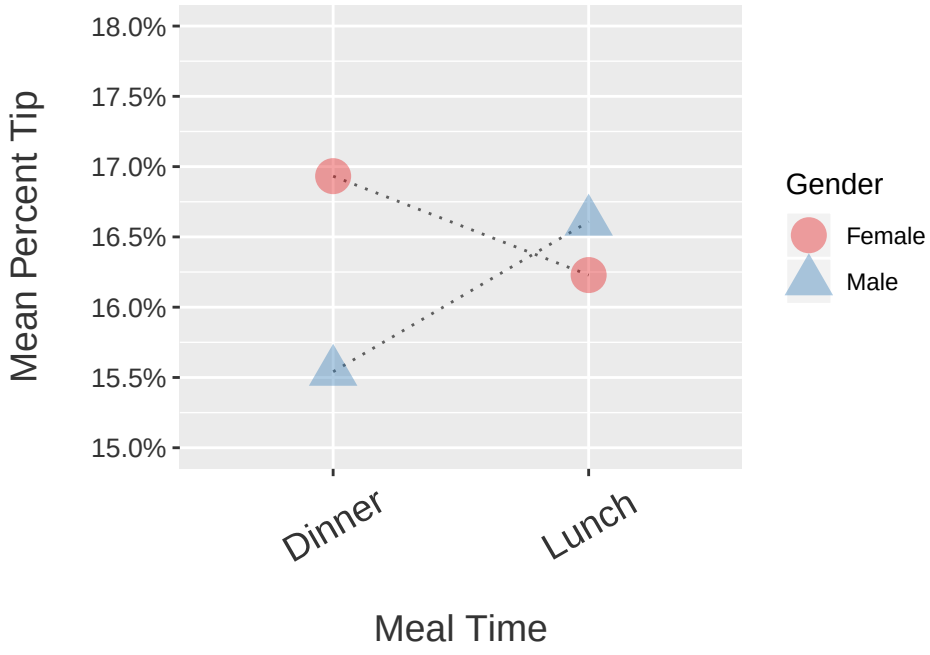
(a) Plot `time` on x-axis, total_bill on the y-axis, colored by `smoker` and shaped by `sex`.



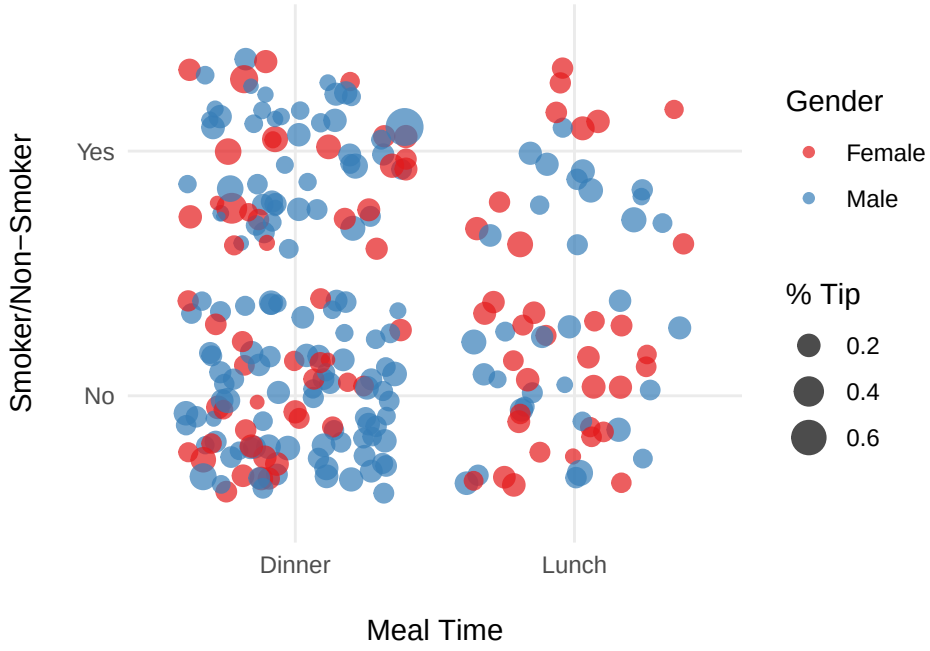
(b) `jitter` the previous plot so points are more visible



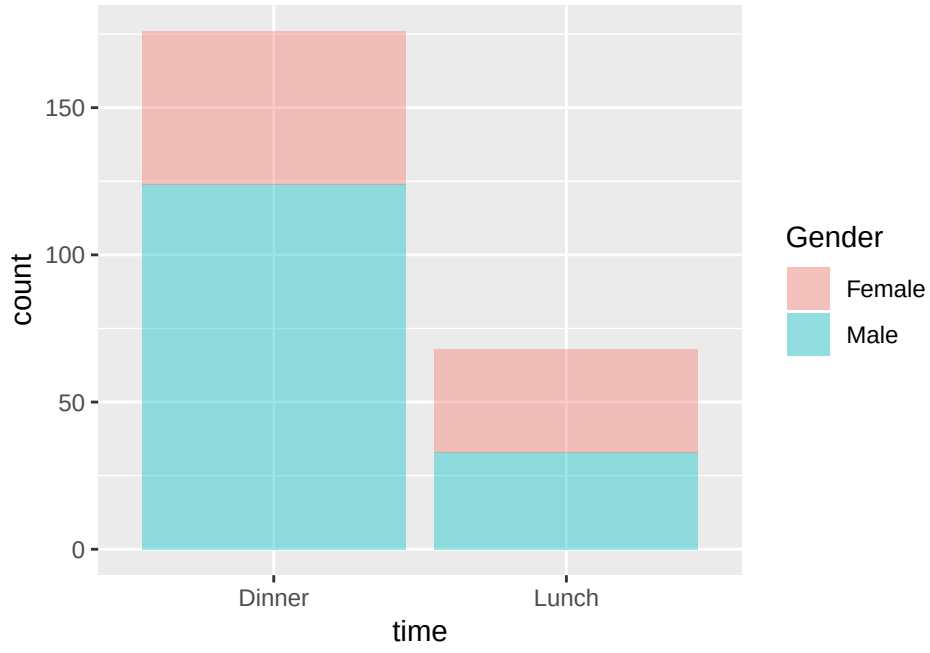
(c) Create a graph with `time` on the x-axis, mean `tip` on the y-axis, shaped and coloured by `sex`.



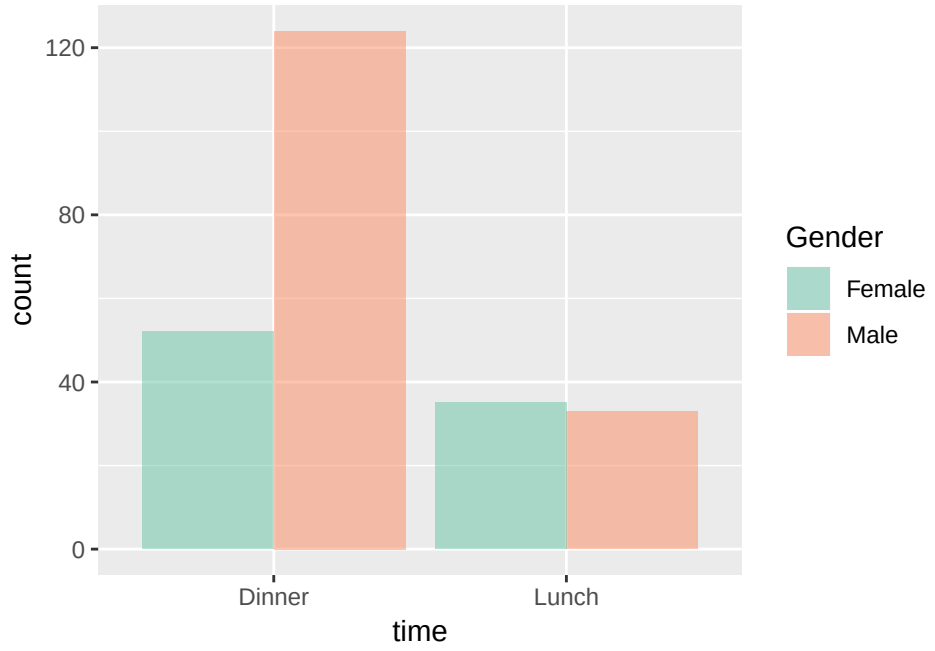
(d) Create a graph with `time` on the x-axis, `smoker` on the y-axis, colored by `sex` and sized by `prctTip`, with a minimalist `theme`.



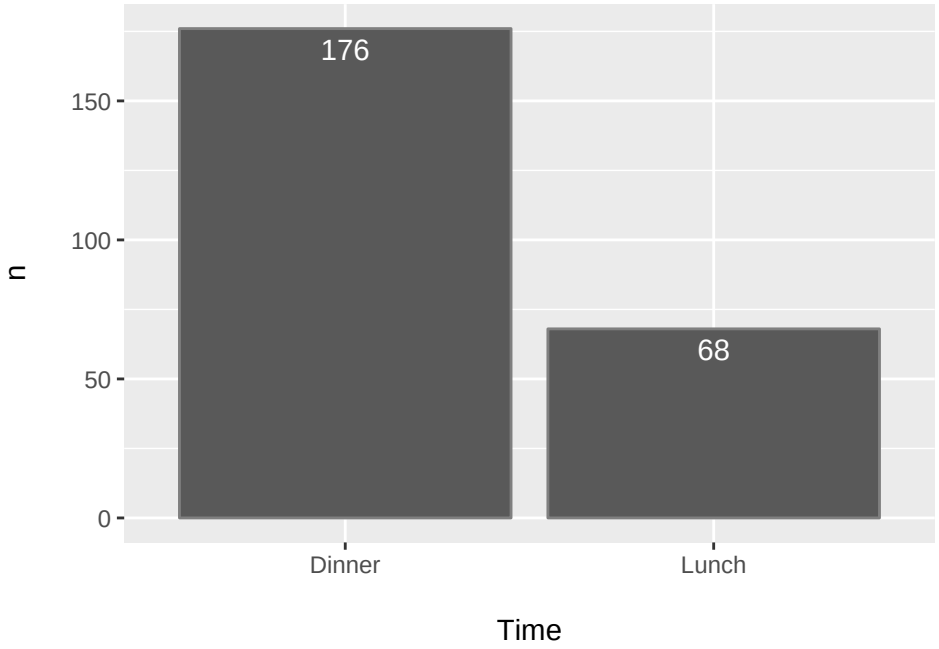
(e) Create a stacked bar graph with `time` on x-axis, `count` on the y-axis, filled by `sex`.



(f) Create a `dodged` bar graph with `time` on x-axis, `count` on the y-axis, filled by `sex`.



(g) Create a bar graph with `time` on x-axis, `count` on the y-axis, and include the values of each bar.



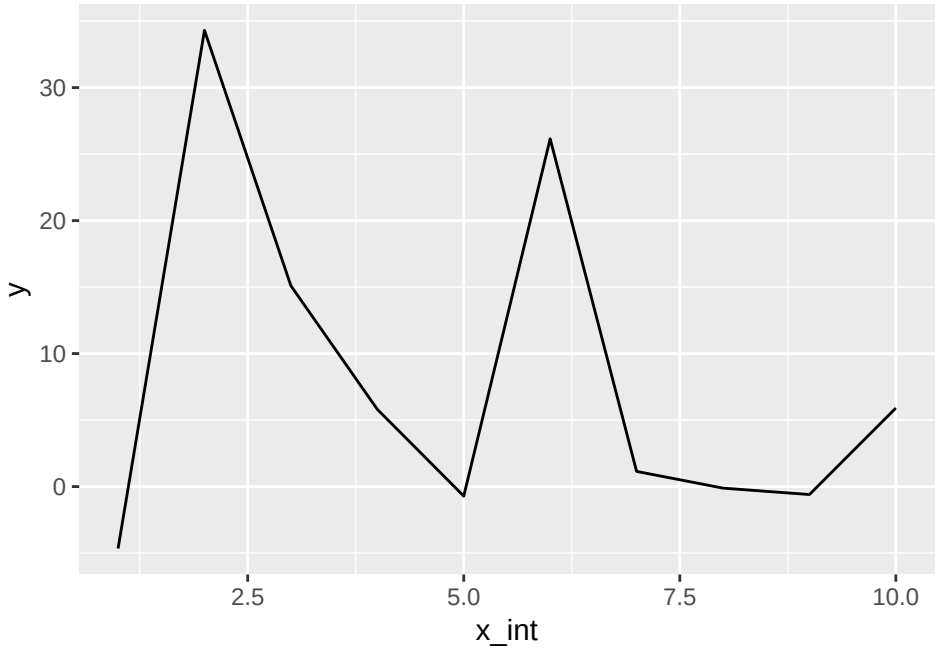
Question 2

Create some fictional data using the following code:

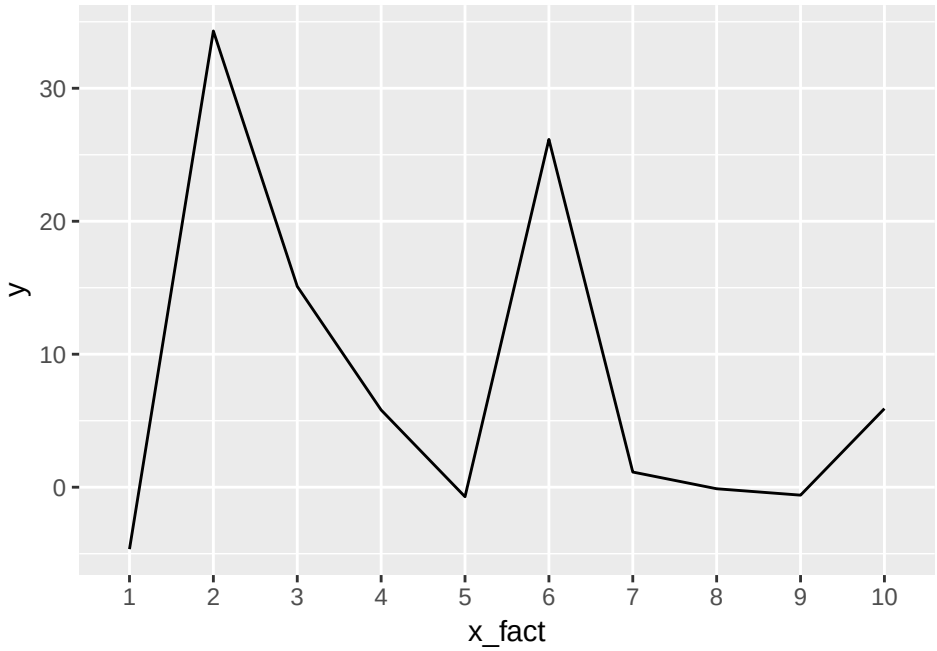
```
set.seed(22)
x_int <- seq(10)
x_fact <- factor(seq(10))
y <- rnorm(10, 2, 13)
myDF <- data_frame(x_int, x_fact, y)
```

Note the `type`s of each of the columns in `myDF`.

(a) Create a line graph with `x_int` on the x-axis and `y` on the y-axis.



(b) Create a line graph with `x_fact` on the x-axis and `y` on the y-axis.

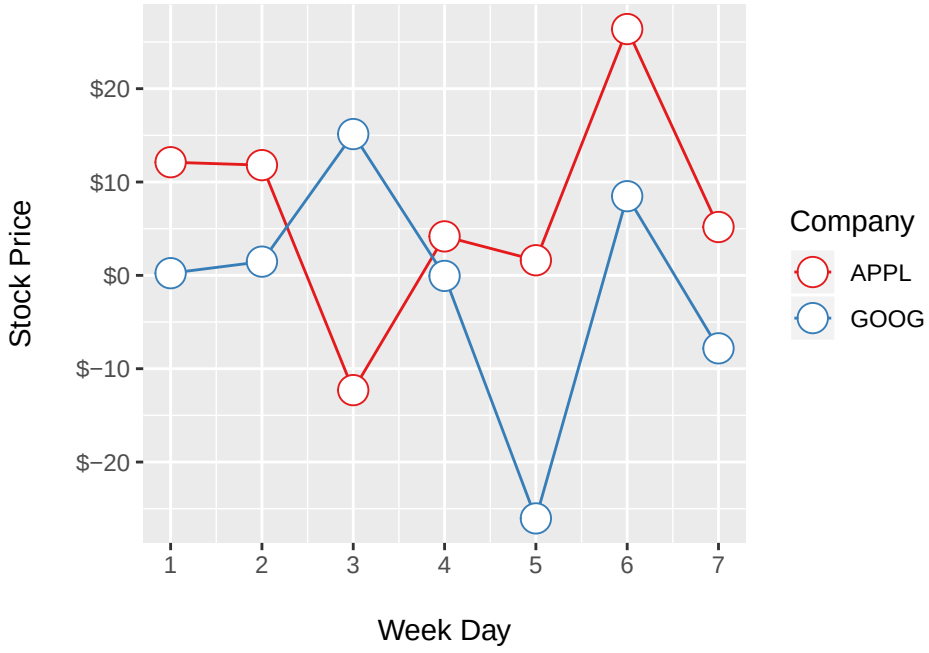


Question 3

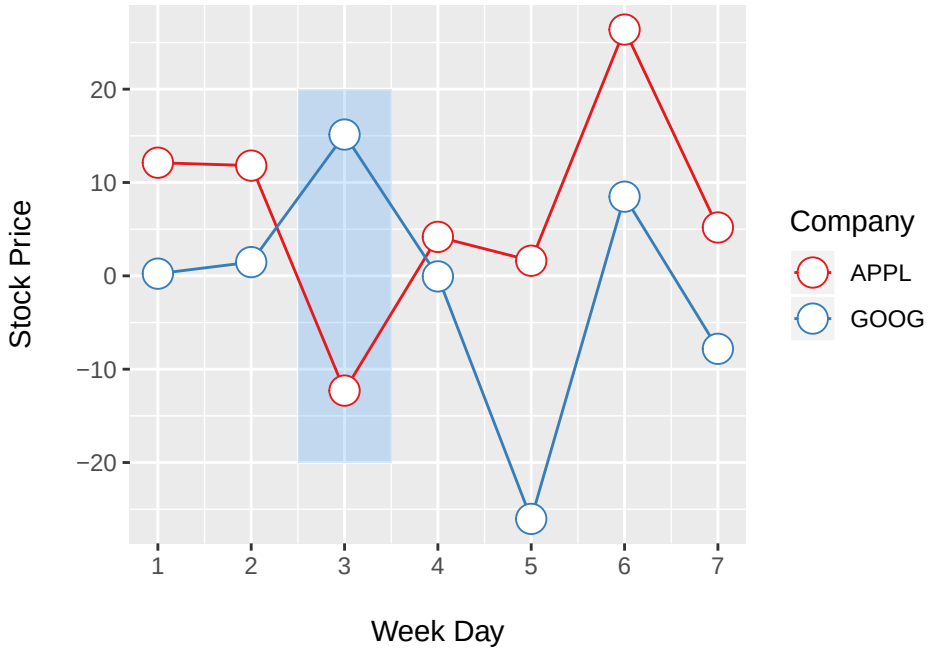
Create some fictional data using the following code:

```
set.seed(33)
day <- rep(seq(7), 2)
stockPrice <- rnorm(14, 2, 13)
company <- c(rep("GOOG", 7), rep("APPL", 7))
myDF <- data_frame(day, stockPrice, company)
```

(a) Create a line and point graph with `day` on the x-axis and `stockPrice` on the y-axis, grouped by `company`.



(b) Add a shaded blue bar to the previous graph, as shown.

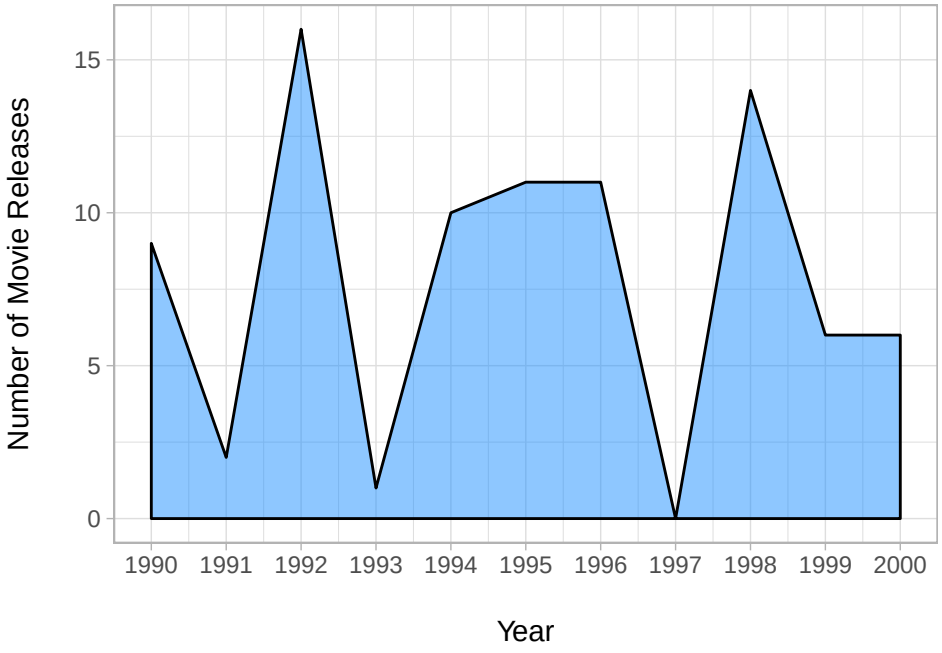


Question 4

Create some fictional data using the following code:

```
set.seed(44)
year <- seq(1990, 2000)
numMovieReleases <- round(abs(rnorm(11, 2, 10)))
genre <- factor(sample(size = 11, c("Horror", "Drama", "SciFi"),
    replace = T))
myDF <- data_frame(year, numMovieReleases, genre)
```

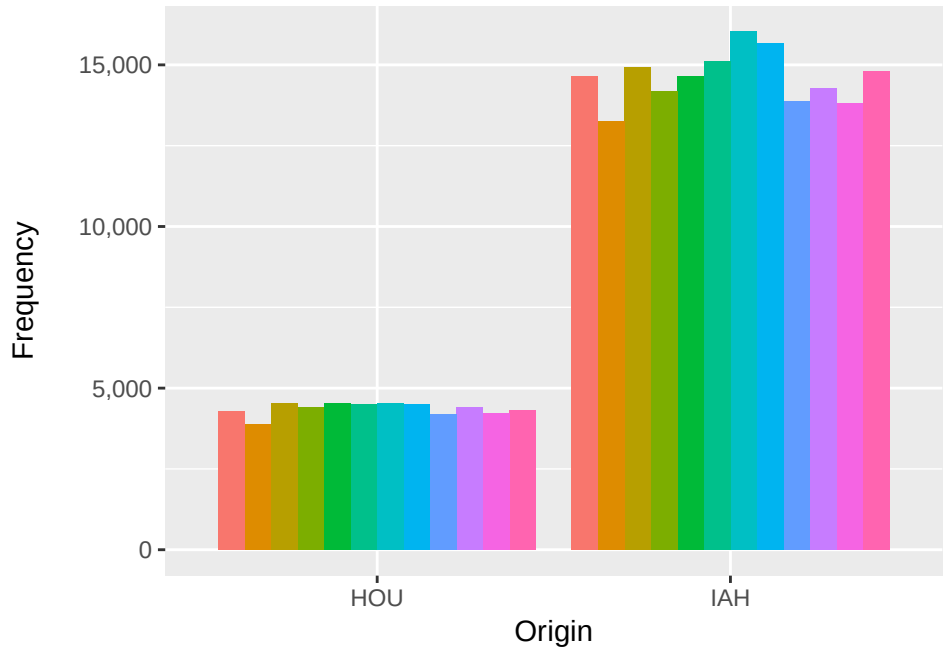
Create an area graph with `year` on the x-axis and `numMovieRelases` on the y-axis.



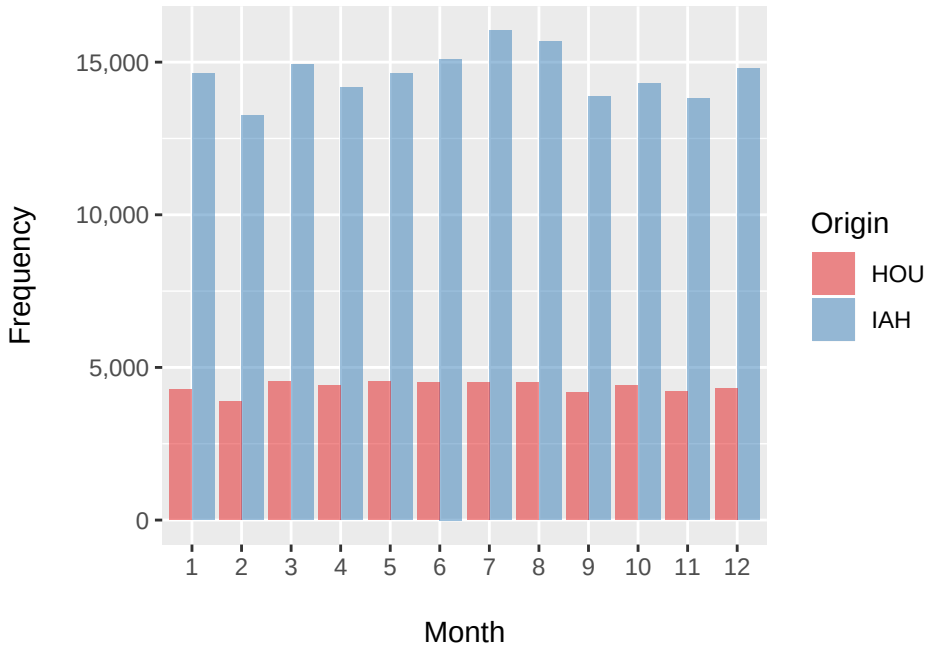
Question 5

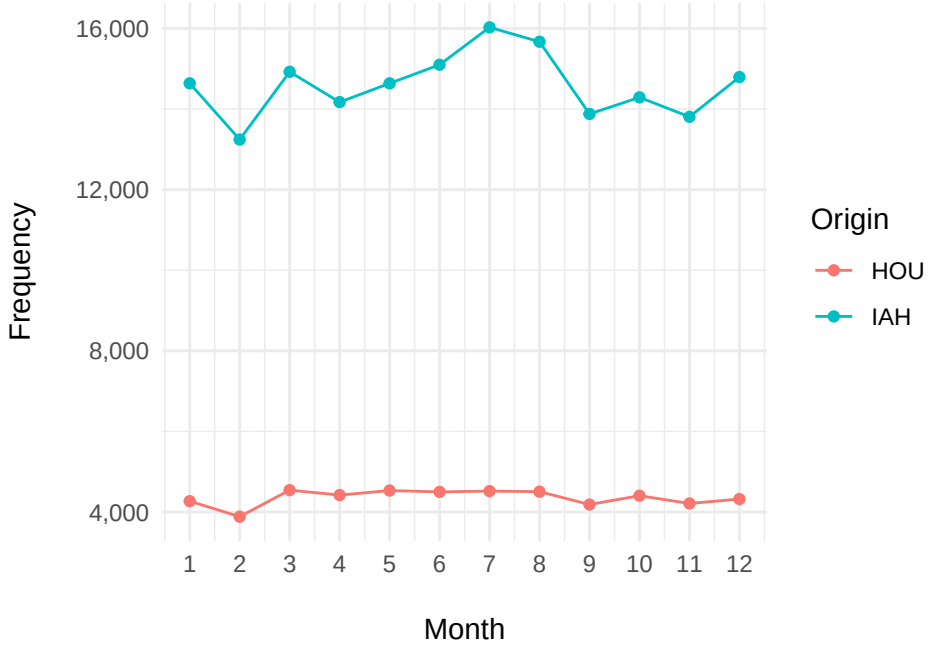
The `hflights` package contains a dataset named `hflights`, which provides information on 227,496 flights in 2011 leaving from Houston-based airports. How many flights departed per month? From IAH? From HOU? Create the following charts.

(a) Create a bar graph with `Origin` on the x-axis, `Frequency` on the y-axis, with `dodge`d bars for each month.



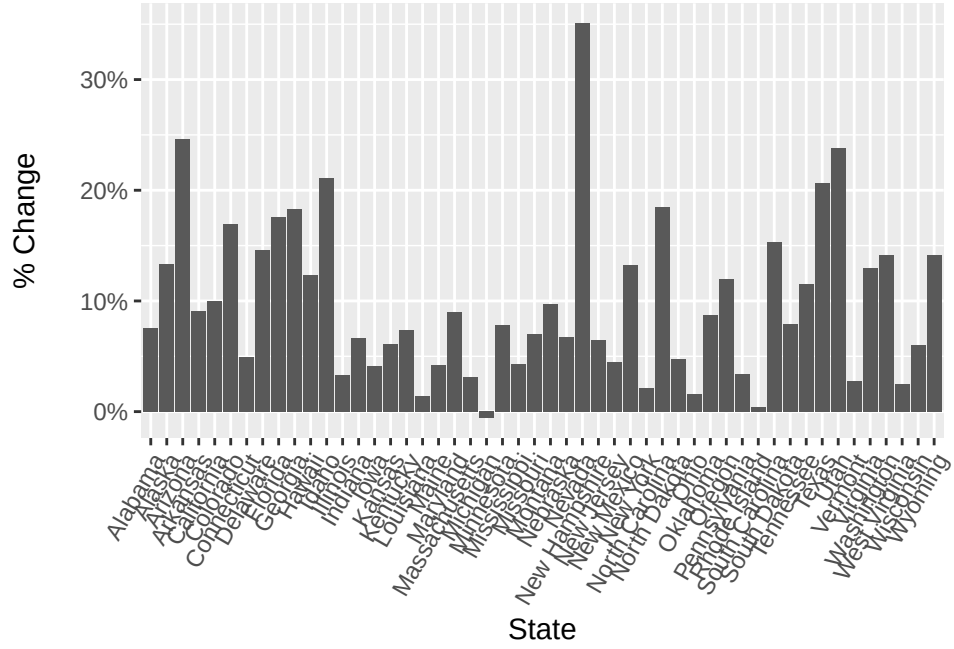
(b) Create a bar graph with `Month` on the x-axis, `Frequency` on the y-axis, with `dodge`d bars for each `Origin`.

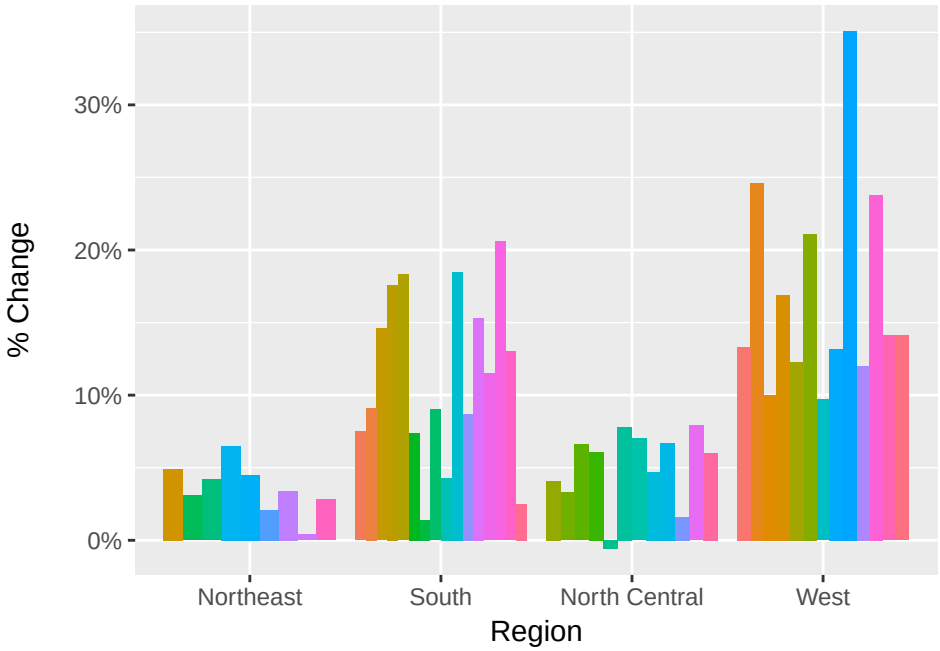

15,000
10,000
5,000
0
Frequency
1
2
3
4
5
6
7
8
9
10
11
12
Month
Origin
HOU
IAH

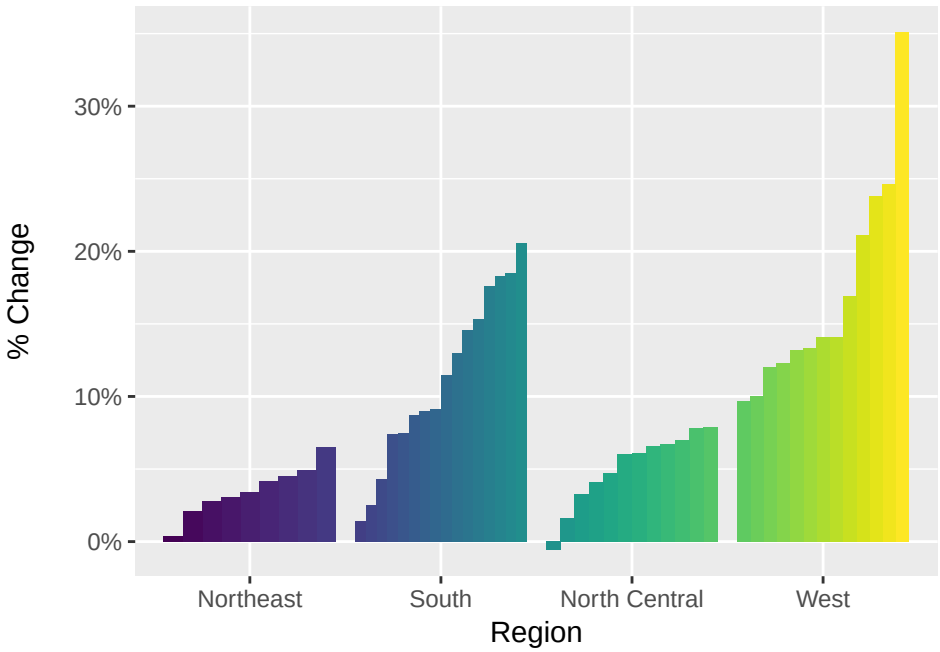

16,000
12,000
8,000
4,000
Frequency
1
2
3
4
5
6
7
8
9
10
11
12
Month
Origin
HOU
IAH

Question 6

The `gcookbook` package contains a dataset named `uspopchange`, which provides information on recent population changes in US States. We are interested in graphing the chnage in population across all states. Create the following graphs.






30%
20%
10%
0%
% Change
Northeast
South
North Central
West
Region