

# BA888 Capstone Project

*Team 1*

*2/13/2020*

```
uber = read_csv("uber_cleandata_with_weather.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   id = col_character(),
##   source = col_character(),
##   destination = col_character(),
##   car_type = col_character(),
##   short_summary = col_character(),
##   icon = col_character()
## )

## See spec(...) for full column specifications.
```

```
glimpse(uber)
```

```
## Observations: 330,568
## Variables: 39
## $ id                <chr> "009e9c53-074d-43cf-aef2-0fbc7a47ed3d"...
## $ hour              <dbl> 22, 10, 19, 23, 0, 19, 20, 3, 13, 10, ...
## $ day               <dbl> 30, 13, 13, 16, 14, 29, 17, 16, 27, 13...
## $ month             <dbl> 11, 12, 12, 12, 12, 11, 12, 12, 11, 12...
## $ source            <chr> "North End", "North End", "North End",...
## $ destination       <chr> "West End", "West End", "West End", "W...
## $ car_type          <chr> "UberXL", "Black", "UberX", "WAV", "Bl...
## $ price             <dbl> 12.0, 16.0, 7.5, 7.5, 26.0, 5.5, 8.5, ...
## $ distance          <dbl> 1.11, 1.11, 1.11, 1.11, 1.11, 1.11, 2...
## $ surge_multiplier  <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ latitude          <dbl> 42.3647, 42.3647, 42.3647, 42.3647, 42...
## $ longitude         <dbl> -71.0542, -71.0542, -71.0542, -71.0542...
## $ temperature       <dbl> 40.13, 20.38, 32.85, 41.29, 31.25, 43...
## $ apparentTemperature <dbl> 38.08, 20.38, 32.85, 36.01, 31.25, 37...
## $ short_summary     <chr> "Overcast", "Clear", "Mostly Cloudy", ...
## $ precipIntensity   <dbl> 0.0000, 0.0000, 0.0000, 0.0567, 0.0000...
## $ precipProbability <dbl> 0.00, 0.00, 0.00, 0.94, 0.00, 0.00, 0...
## $ humidity          <dbl> 0.60, 0.66, 0.56, 0.86, 0.64, 0.52, 0...
## $ windSpeed         <dbl> 3.38, 2.94, 2.65, 8.30, 2.62, 12.13, 6...
## $ windGust          <dbl> 3.99, 3.22, 3.83, 8.30, 3.54, 19.97, 8...
## $ visibility        <dbl> 9.833, 9.831, 9.959, 4.054, 10.000, 9...
## $ temperatureHigh   <dbl> 42.52, 33.83, 33.83, 43.83, 33.83, 44...
## $ temperatureLow    <dbl> 31.71, 27.27, 27.27, 34.25, 27.27, 28...
## $ apparentTemperatureHigh <dbl> 40.53, 32.85, 32.85, 38.38, 32.85, 38...
## $ apparentTemperatureLow <dbl> 28.06, 24.61, 24.61, 28.30, 24.61, 26...
## $ icon              <chr> "cloudy", "clear", "partly-cloudy", "r...
## $ dewPoint          <dbl> 27.31, 10.87, 18.66, 37.56, 20.53, 26...
```

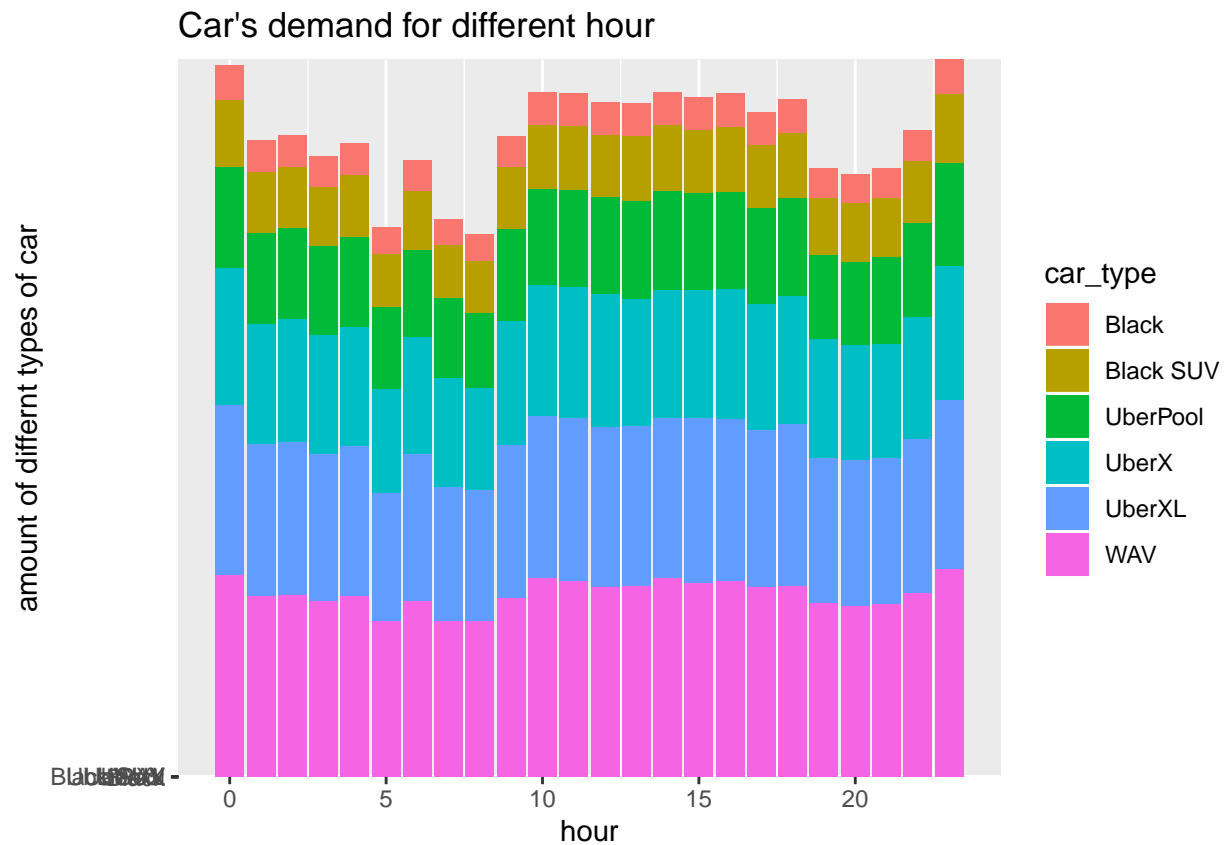
```
## $ pressure          <dbl> 1017.16, 1031.51, 1033.65, 1012.72, 10...
## $ windBearing       <dbl> 281, 2, 76, 57, 173, 313, 286, 38, 227...
## $ cloudCover        <dbl> 1.00, 0.03, 0.64, 1.00, 0.91, 0.53, 0....
## $ uvIndex           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ visibility.1      <dbl> 9.833, 9.831, 9.959, 4.054, 10.000, 9....
## $ ozone             <dbl> 281.8, 327.3, 330.8, 325.3, 326.7, 309...
## $ moonPhase         <dbl> 0.79, 0.21, 0.21, 0.30, 0.21, 0.75, 0....
## $ precipIntensityMax <dbl> 0.0003, 0.0001, 0.0001, 0.1252, 0.0001...
## $ temperatureMin    <dbl> 28.79, 18.29, 18.29, 39.22, 18.29, 35....
## $ temperatureMax    <dbl> 42.52, 33.83, 33.83, 43.83, 33.83, 44....
## $ apparentTemperatureMin <dbl> 26.41, 13.79, 13.79, 33.98, 13.79, 31....
## $ apparentTemperatureMax <dbl> 40.53, 32.85, 32.85, 38.38, 32.85, 38....
```

```
head(uber, 10)
```

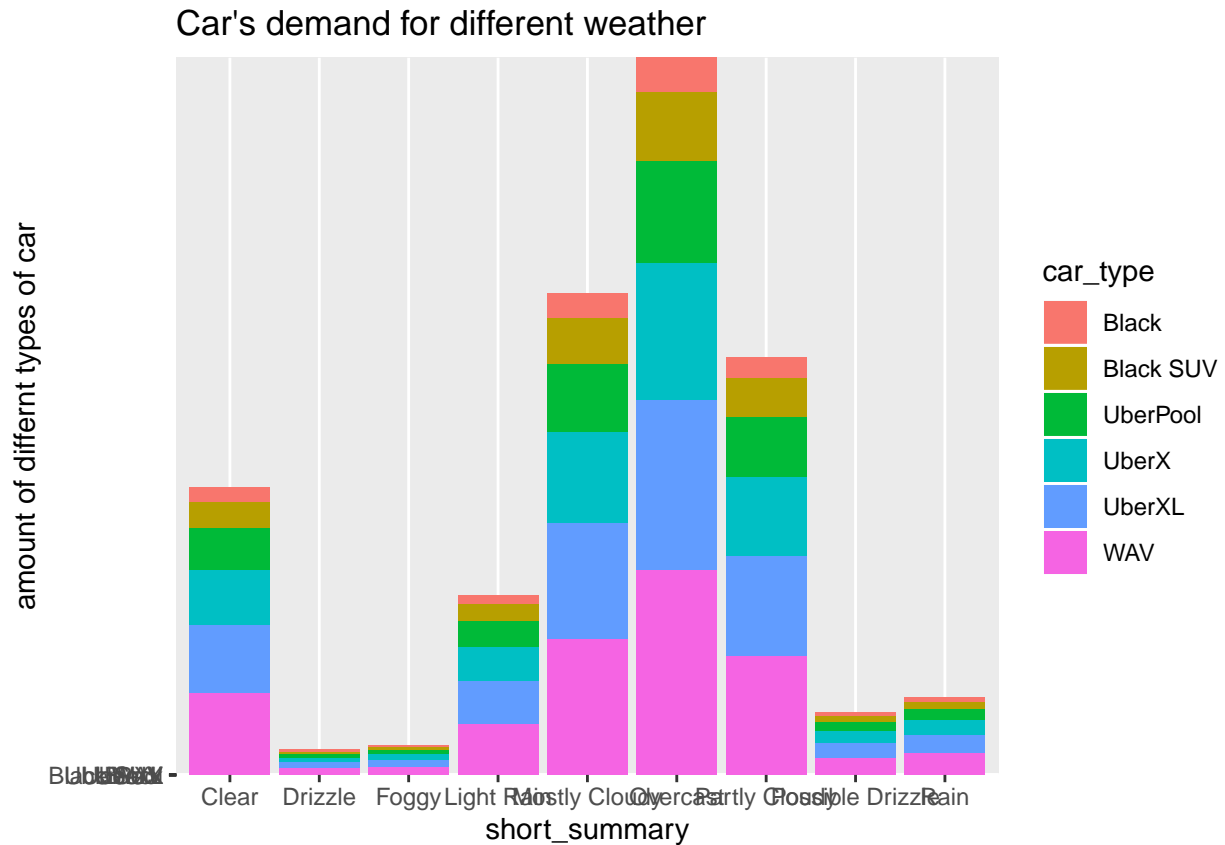
```
## # A tibble: 10 x 39
##   id      hour  day month source destination car_type price distance
##   <chr> <dbl> <dbl> <dbl> <chr> <chr>      <chr>    <dbl>    <dbl>
## 1 009e~    22    30    11 North~ West End   UberXL     12      1.11
## 2 23f1~    10    13    12 North~ West End   Black      16      1.11
## 3 3575~    19    13    12 North~ West End   UberX       7.5     1.11
## 4 50ef~    23    16    12 North~ West End   WAV         7.5     1.11
## 5 91c4~     0    14    12 North~ West End   Black S~    26      1.11
## 6 e219~    19    29    11 North~ West End   UberPool    5.5     1.11
## 7 1d45~    20    17    12 Beaco~ South Stat~ WAV         8.5     2.48
## 8 1f64~     3    16    12 Beaco~ South Stat~ UberXL     15     2.48
## 9 2ca4~    13    27    11 Beaco~ South Stat~ Black     20.5    2.48
## 10 4149~    10    13    12 Beaco~ South Stat~ UberX      8.5     2.48
## # ... with 30 more variables: surge_multiplier <dbl>, latitude <dbl>,
## # longitude <dbl>, temperature <dbl>, apparentTemperature <dbl>,
## # short_summary <chr>, precipIntensity <dbl>, precipProbability <dbl>,
## # humidity <dbl>, windSpeed <dbl>, windGust <dbl>, visibility <dbl>,
## # temperatureHigh <dbl>, temperatureLow <dbl>,
## # apparentTemperatureHigh <dbl>, apparentTemperatureLow <dbl>,
## # icon <chr>, dewPoint <dbl>, pressure <dbl>, windBearing <dbl>,
## # cloudCover <dbl>, uvIndex <dbl>, visibility.1 <dbl>, ozone <dbl>,
## # moonPhase <dbl>, precipIntensityMax <dbl>, temperatureMin <dbl>,
## # temperatureMax <dbl>, apparentTemperatureMin <dbl>,
## # apparentTemperatureMax <dbl>
```

Team 1 discusses about how location and weather reflects uber price and try to find how what we found will affect some industry like hotel, which have connections with ordering service. Our data has 39 variables and 330568 rows, which have four different categories data: time, location, price and weather information.

```
ggplot(uber)+
  geom_bar(aes(x=hour,y=car_type,fill=car_type),stat='identity')+
  ylab("amount of different types of car")+
  ggtitle("Car's demand for different hour")
```

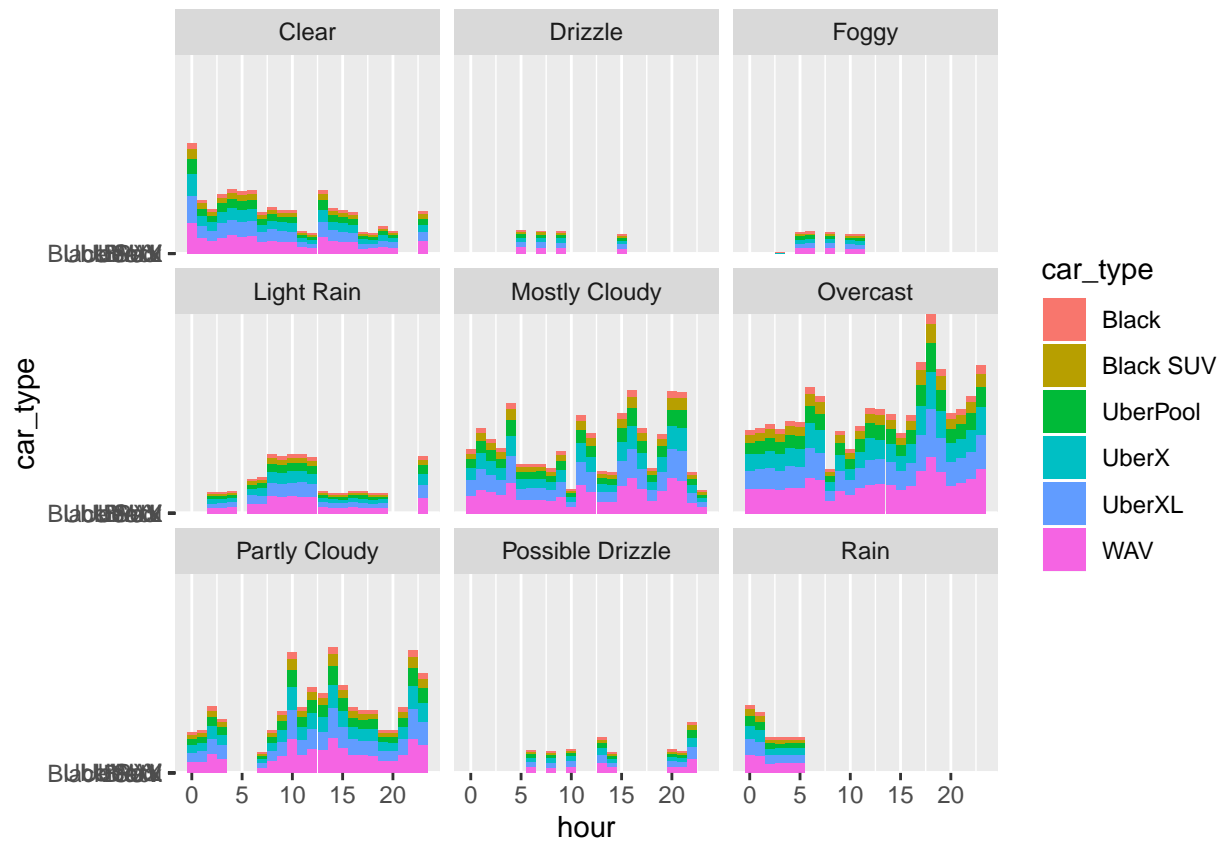


```
ggplot(uber,aes(x=short_summary,y=car_type,fill=car_type))+
  geom_bar(stat='identity')+
  ylab("amount of differnt types of car")+
  ggtitle("Car's demand for different weather")
```



From first chart we can see highest car demand during 11pm and 12am, and the second higher demand for cars is around 10am and 5pm. The second chart we can see the highest car demand on weather on partly cloudy and cloudy, and people have too much demand on good weather of rainy day. We can also see most popular car is WAV, which uber drivers provide affordable rides in wheelchair. Even though most user don't know whether their car provide wheelchair, but most of them order cars with this service. The second most popular cartype is UberXL.

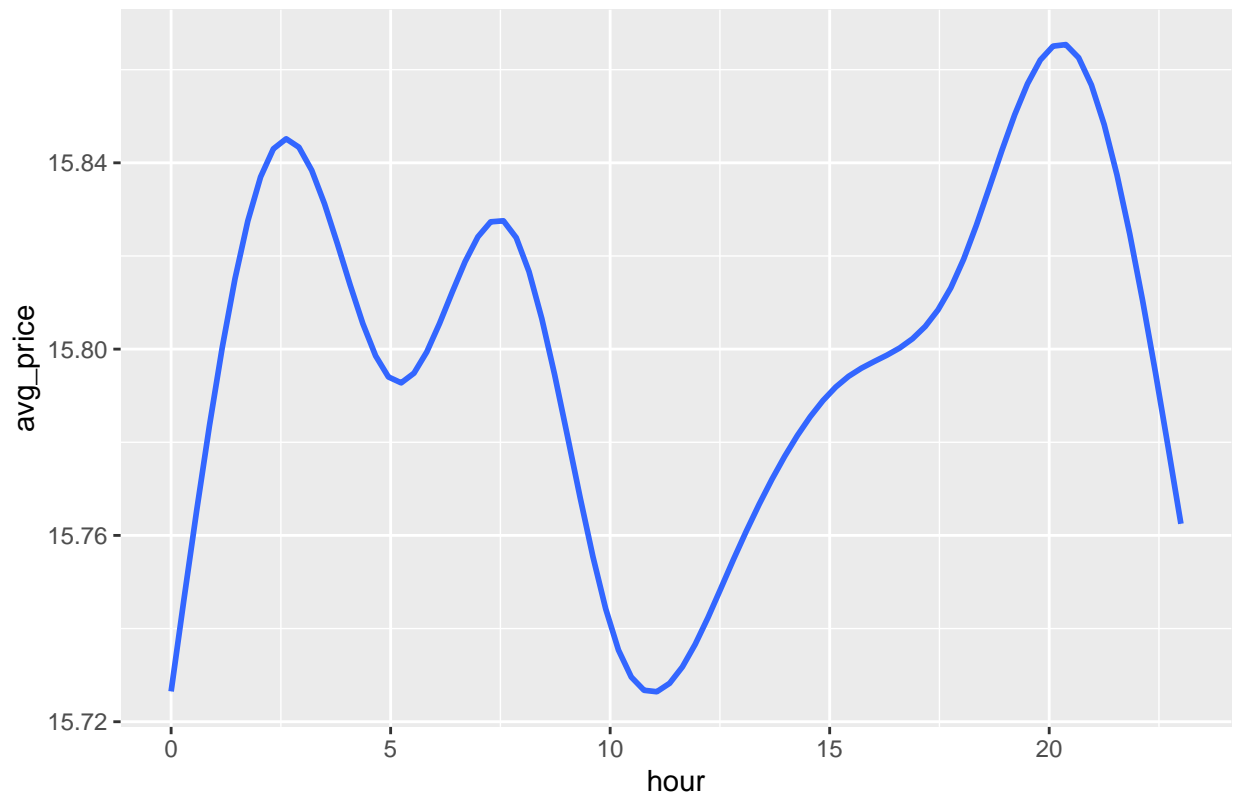
```
ggplot(uber,aes(x=hour,y=car_type,fill=car_type))+
  geom_bar(stat='identity')+
  facet_wrap(~short_summary)
```



```
## price and car amount change during 24h
uber %>% select(price, hour, short_summary) %>% group_by(hour) %>%
  mutate(avg_price = mean(price)) %>%
  ggplot() +
  labs(title = "Price change during the day") +
  geom_smooth(aes(x = hour, y = avg_price), na.rm = FALSE, se = TRUE)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

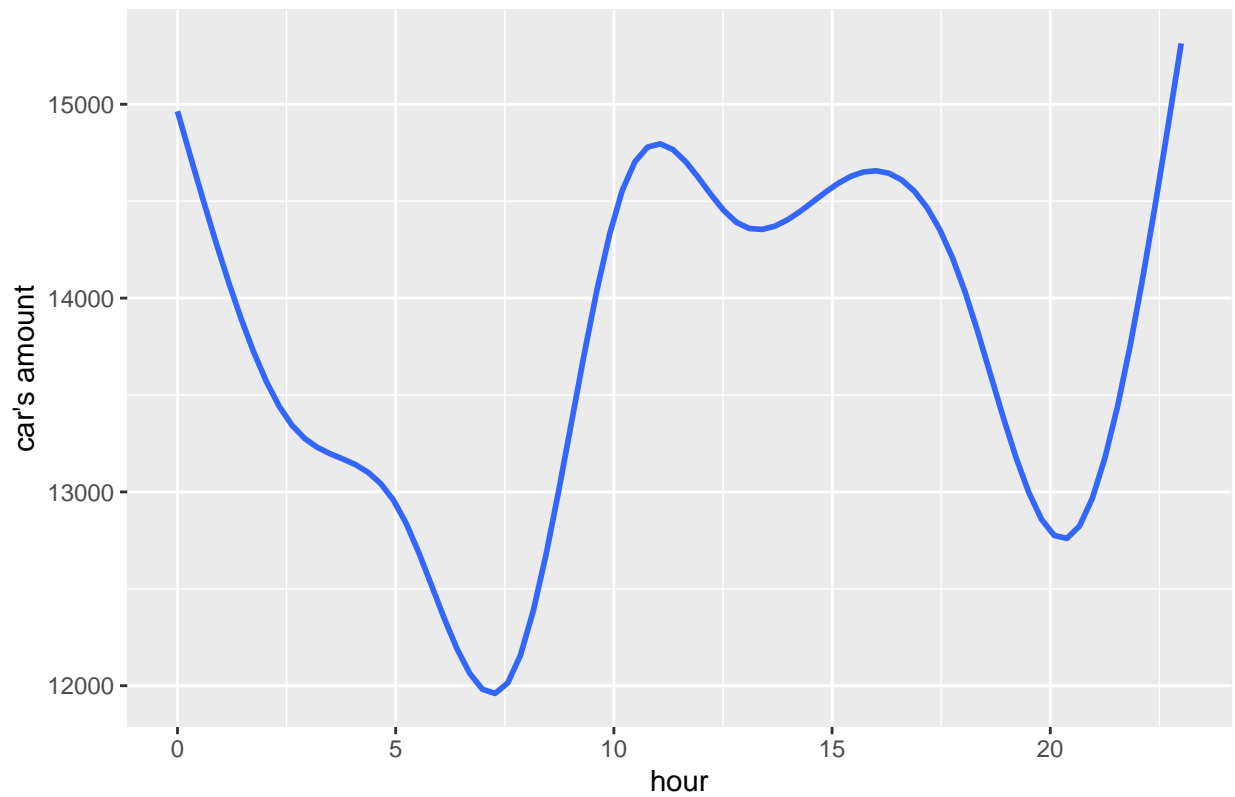
Price change during the day



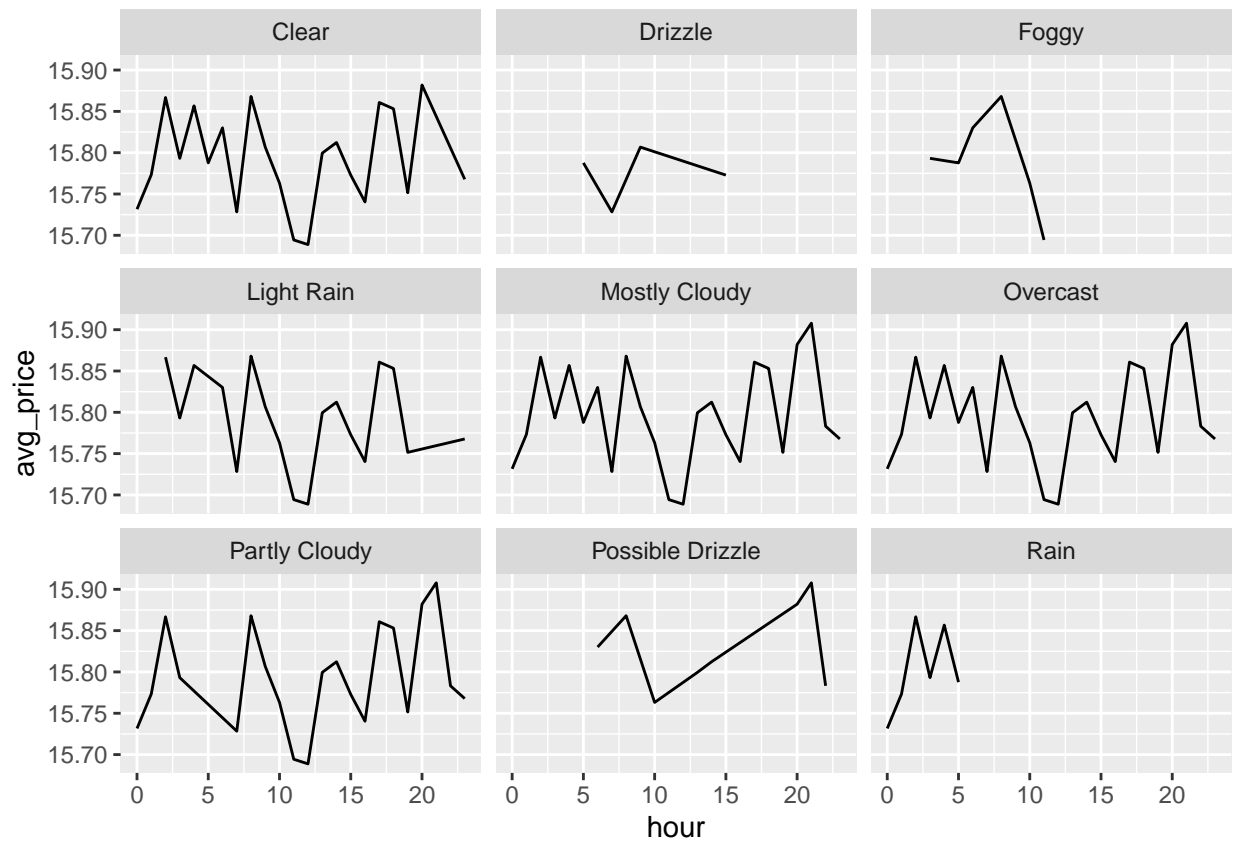
```
uber %>% select(hour, car_type, short_summary) %>% group_by(hour) %>%  
  mutate(count=n()) %>%  
  ggplot() +  
  geom_smooth(aes(x=hour, y=count), na.rm = FALSE, se = TRUE) +  
  labs(y="car's amount", title = "car's requirement change during the day")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

car's requirement change during the day



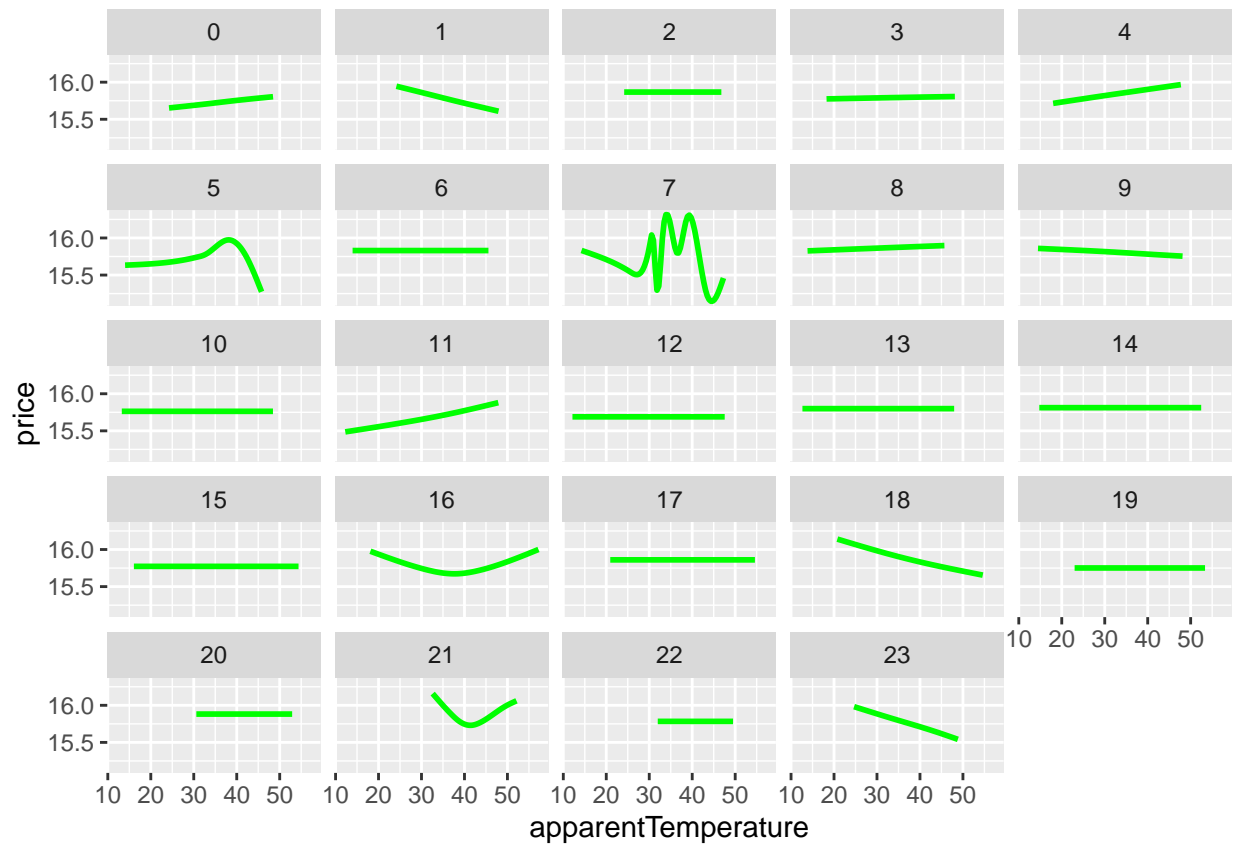
```
uber %>% select(price, hour, short_summary) %>% group_by(hour) %>%  
  mutate(avg_price=mean(price)) %>%  
  ggplot()+  
  geom_line(aes(x=hour, y=avg_price))+  
  facet_wrap(~short_summary)
```



```
uber_temperature = uber %>% select(hour,car_type,price,temperature,apparentTemperature,temperatureHigh,
uber_temperature %>%
  ggplot() +
  geom_smooth(aes(x=apparentTemperature,y=price),se=F,col="green")+
  facet_wrap(~hour)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```





```
uber_temperature %>%
  ggplot() +
  geom_smooth(aes(x=temperature,y=price),se=F,col="red")+
  facet_wrap(~hour)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

