Data Viz Homework

Annop Techa

Install Packages

```
install.packages(c("ggplot2","tidyverse","patchwork",
          "lubridate"))
```

```
## Installing packages into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
## also installing the dependencies 'sys', 'bit', 'ps', 'colorspace', 'rematch', 'askpass', 'bit64', 'pi
```

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2
## --
```

```
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.4     v forcats 1.0.0
## v purrr   1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(patchwork)
library(lubridate)
```

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

Prep Data

```
glimpse(diamonds)
```

```
## Rows: 53,940
## Columns: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0.~
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ver~
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I,~
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS1, ~
## $ depth   <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4, 64~
```

```
## $ table   <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62, 58~
## $ price   <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340, 34~
## $ x       <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00, 4.~
## $ y       <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05, 4.~
## $ z       <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39, 2.~
```

```
dm <- diamonds
```

Chart 1 Count by diamond grade

```
ggplot(dm,aes(cut,fill = color)) +
  geom_bar() +
  labs(title = "Count diamonds grade ",
  x="Cut",
  y="Count",
  caption = "Sorce : daimonds")  +
  theme_minimal()
```

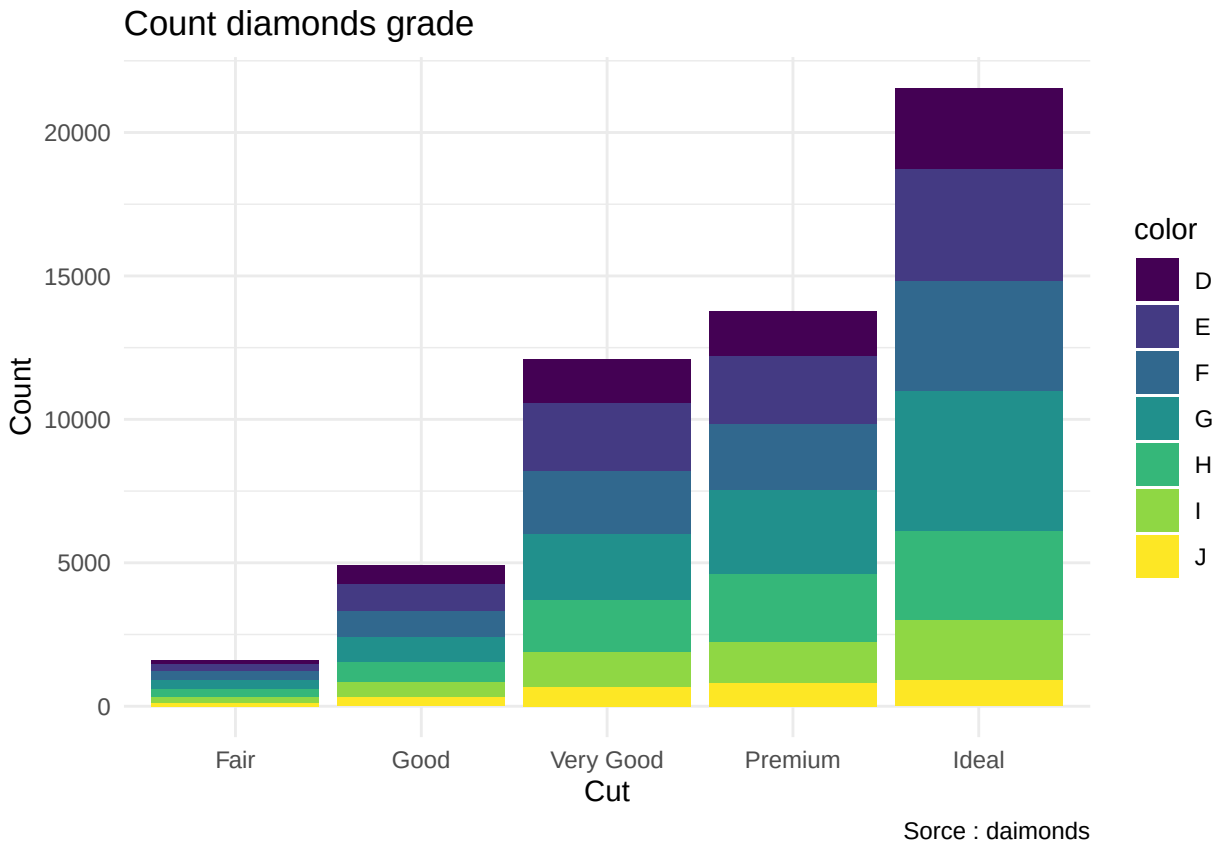


Chart 2 Relationship between Carat & Price

```
set.seed(20)
ggplot(sample_n(dm,500),aes(carat,price,color = color)) +
  geom_point(alpha=0.7,size=0.7) +
  geom_smooth(se=F) +
  theme_minimal()+
```

```
  labs(title = "Relationship between carat & price",
  x="Carat",
  y="Price (USD)",
  caption = "Sorce : daimonds")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

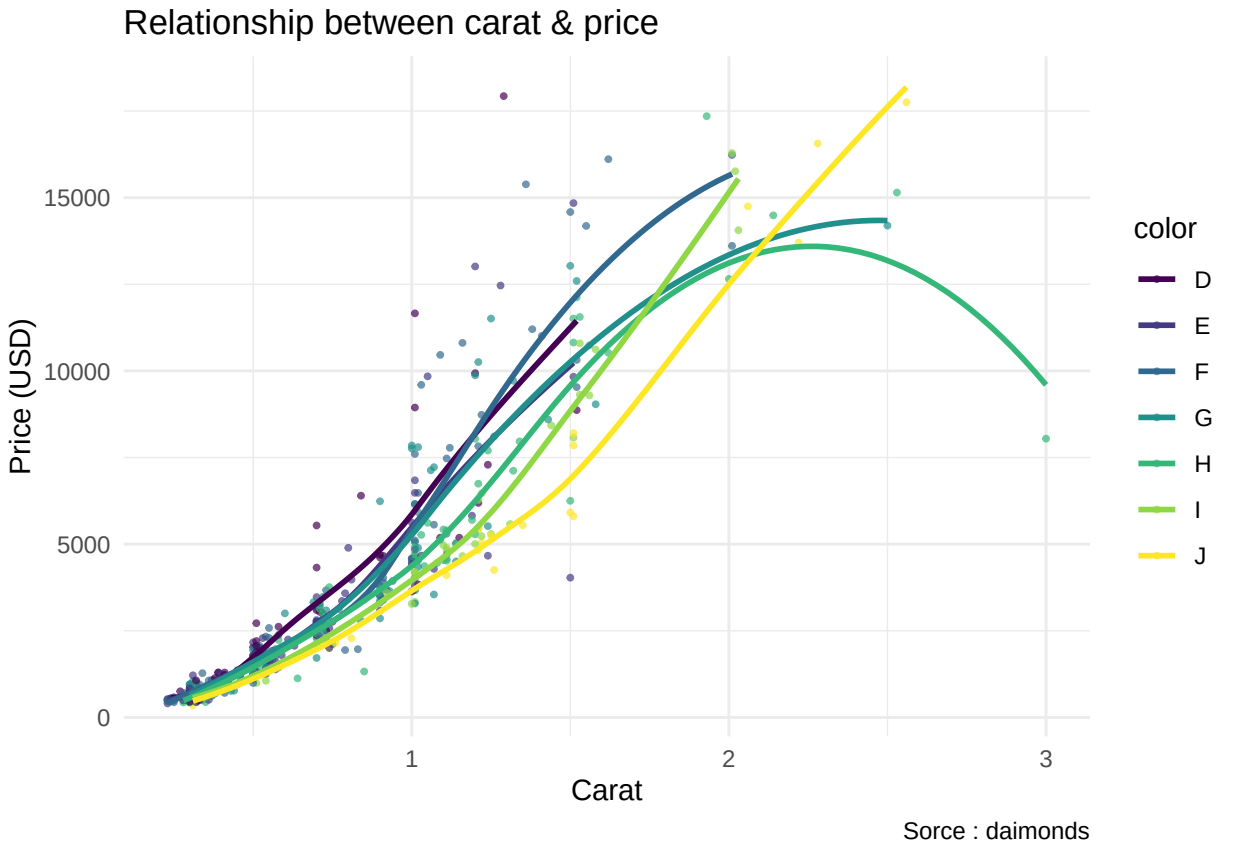


Chart 3 Relationship between Cut & Price

```
set.seed(25)
  ggplot(sample_n(dm,1000),aes(cut,price)) +
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +
  coord_flip() +
  theme_bw() +
  labs(title = "Relationship between cut & price",
  x="Cut",
  y="Price (USD)",
  caption = "Sorce : daimonds")
```

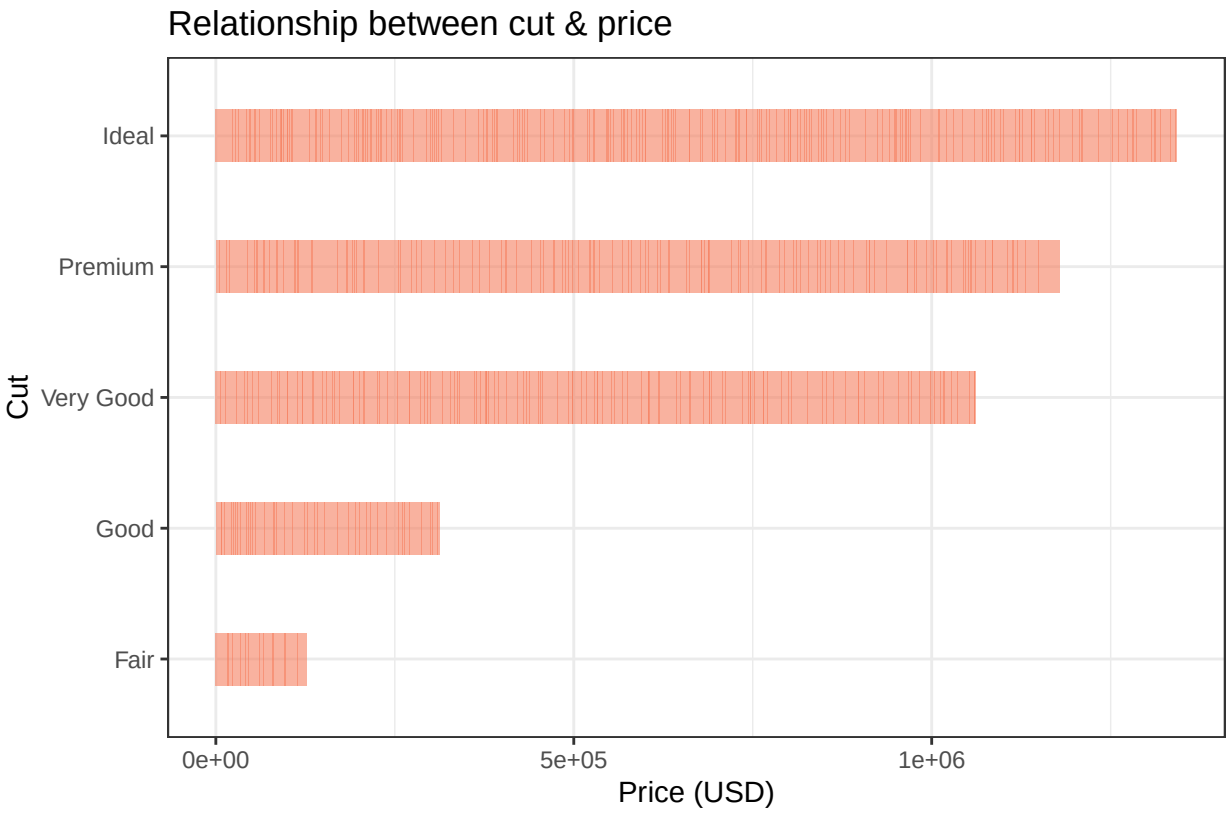
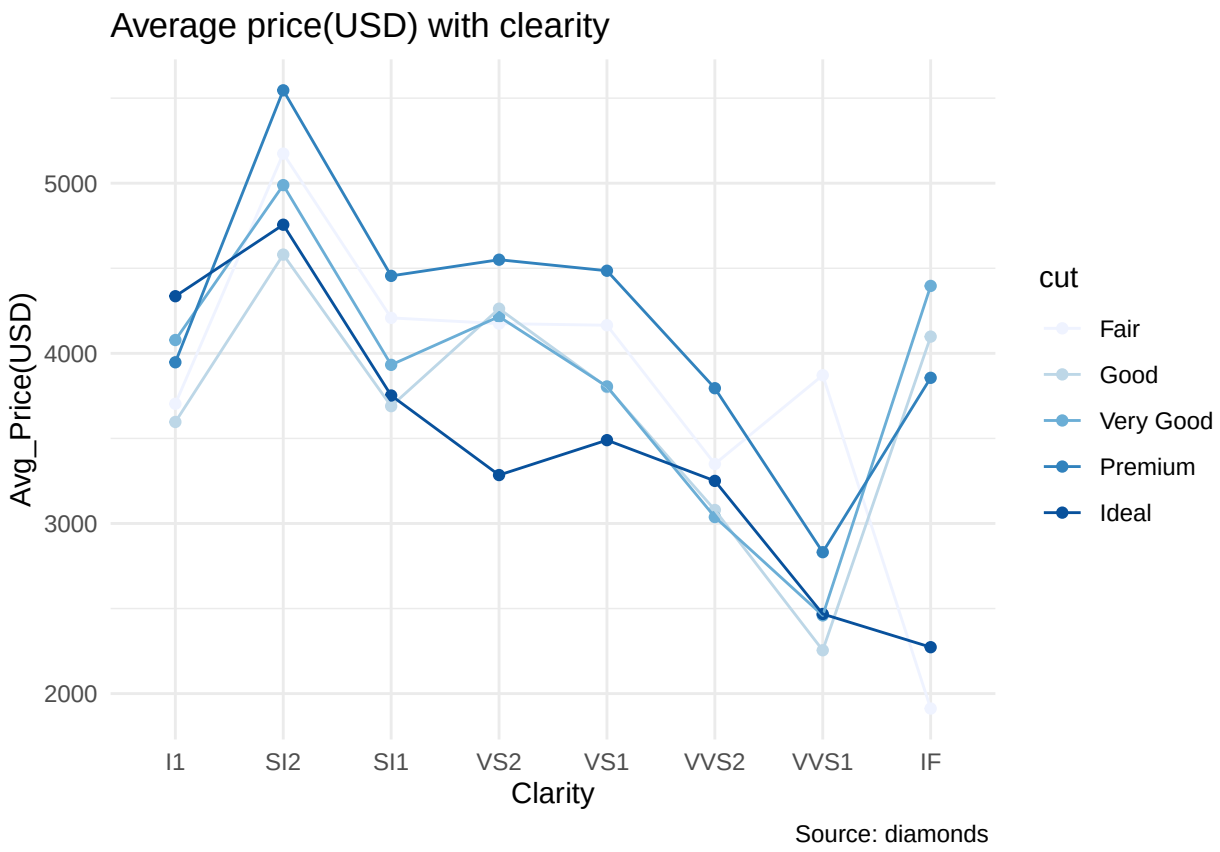


Chart 4 Relationship between Average Price with clearity

```
dm %>%
  group_by(clarity,cut) %>%
  summarise(avg_price = mean(price)) %>%
  ggplot(aes(clarity, avg_price,group = cut, col = cut))+
  geom_point()+
  geom_line()+
  theme_minimal()+
  scale_color_brewer(type ="seq" , palette = 1 ) +
  labs(title = "Average price(USD) with clearity",
  x = "Clarity",
  y = "Avg_Price(USD)",
  caption = "Source: diamonds ")
```

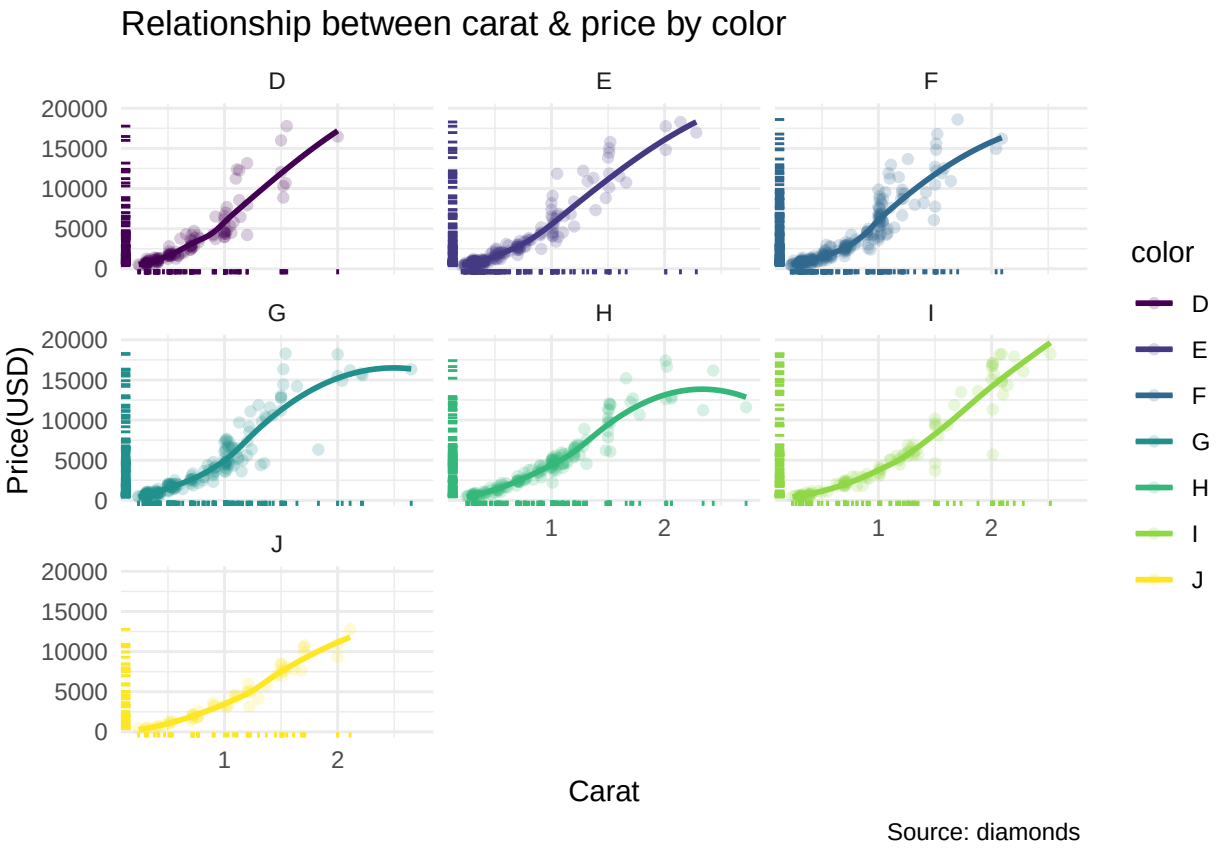
```
## `summarise()` has grouped output by 'clarity'. You can override using the
## `.groups` argument.
```



5 Relationship between Carat & Price by color

```
set.seed(25)
ggplot(sample_n(dm,1000),aes(carat,price,color=color)) +
  geom_point(alpha =0.2) +
  geom_smooth(se=F) +
  geom_rug() +
  theme_minimal() +
  facet_wrap(~color,ncol=3) +
  labs(title = "Relationship between carat & price by color",
  x = "Carat",
  y = "Price(USD)",
  caption = "Source: diamonds ")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```


Relationship between carat & price by color
D
E
F
G
H
I
J
20000
15000
10000
5000
0
1
2
Price(USD)
Carat
color
D
E
F
G
H
I
J
Source: diamonds