Data Visualization Homework

Chitsanupong_O.

install_Packages

install.packages(c("tidyverse","lubridate","ggthemes","ggplot2","dplyr"))

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

```
library(ggthemes)
library(ggplot2)
library(dplyr)
```

Review 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.~
```

Chart 1: Diamonds Carat and Price

```
set.seed(81)
ggplot(sample_n(diamonds, 3000), aes(carat, price, col = color)) +
geom_point() +
geom_smooth(method = "lm",color = "Blue") +
labs(title = "Relationship between Carat and Price",
x = "Carat",
y = "Price (USD)",
caption = "Source: Diamonds dataset") +
theme_minimal() +
scale_color_brewer(palette = "RdPu") +
facet_wrap(~color, ncol = 4)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

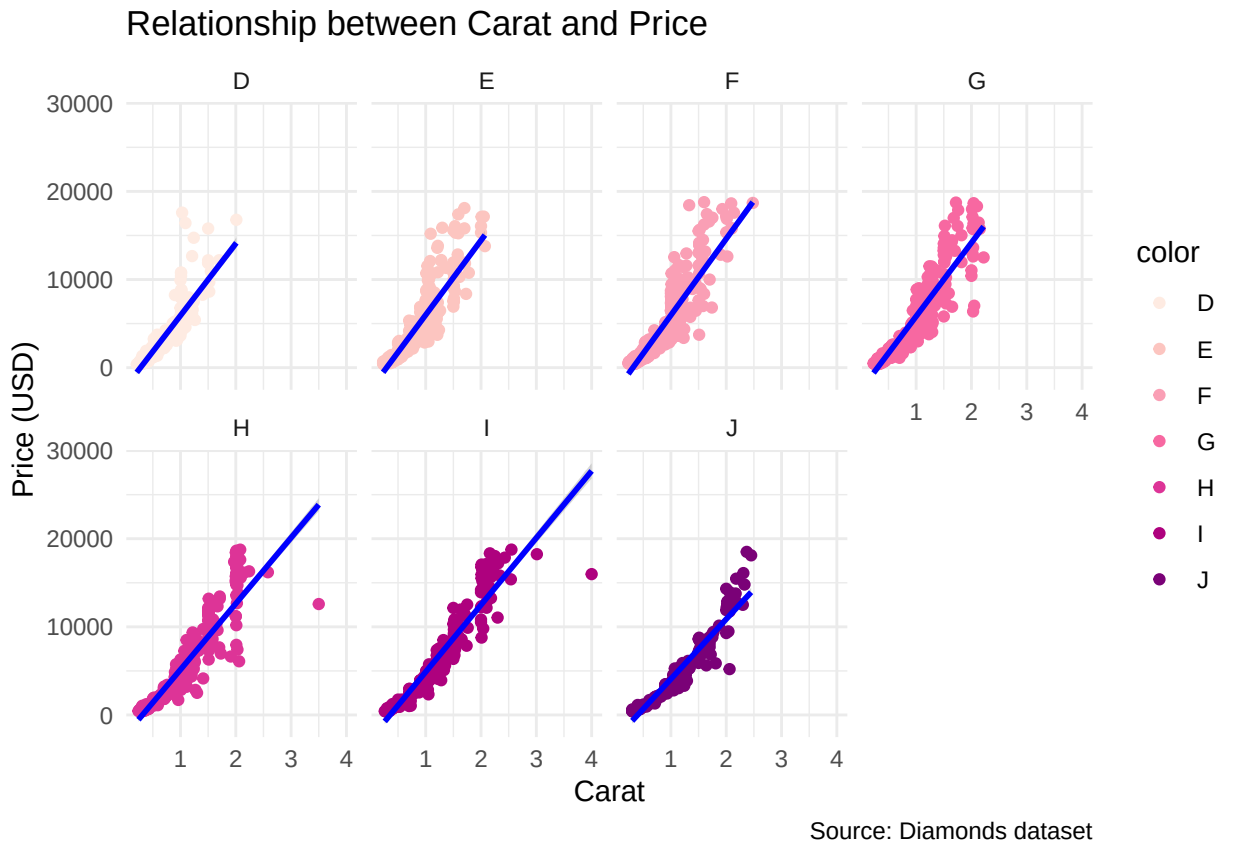


Chart 2: Carat and Frequency

```
set.seed(81)
ggplot(sample_n(diamonds, 3000), aes(x = carat, fill = cut)) +
geom_histogram() +
labs(title = "Relationship between Carat and Frequency",
x = "Carat",
y = "Frequency",
caption = "Source: Diamonds dataset") +
```

```
theme_minimal() +
scale_fill_brewer(type = "seq", palette = "Purples")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

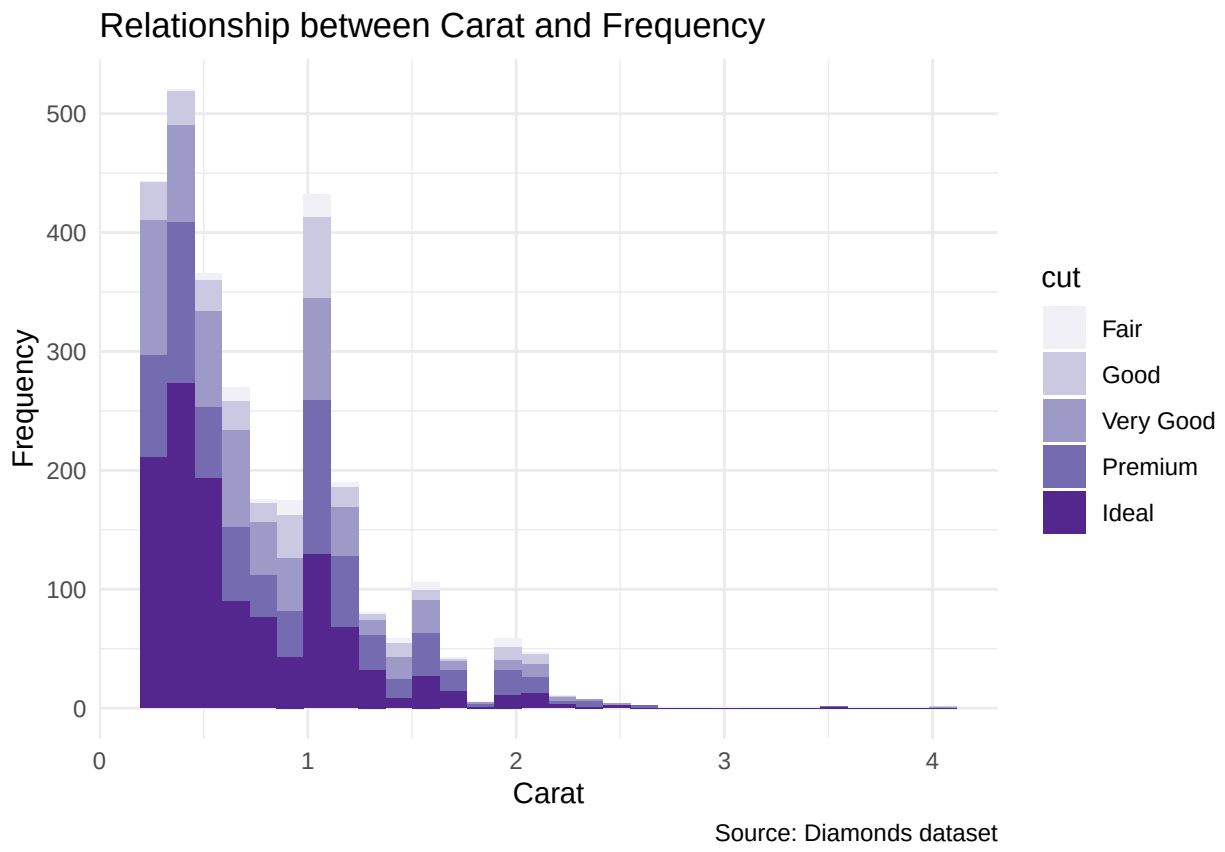


Chart 3: Carat and Clarity

Relationship between Diamonds carat and clarity

```
set.seed(81)
ggplot(sample_n(diamonds,3000), aes(cut, fill=clarity)) +
geom_bar(position = "fill") +
scale_fill_brewer(type = "seq", palette = "Purples") +
labs(title = "Relationship between Cut and Clarity",
caption="Source: Diamonds dataset") +
theme_minimal()
```

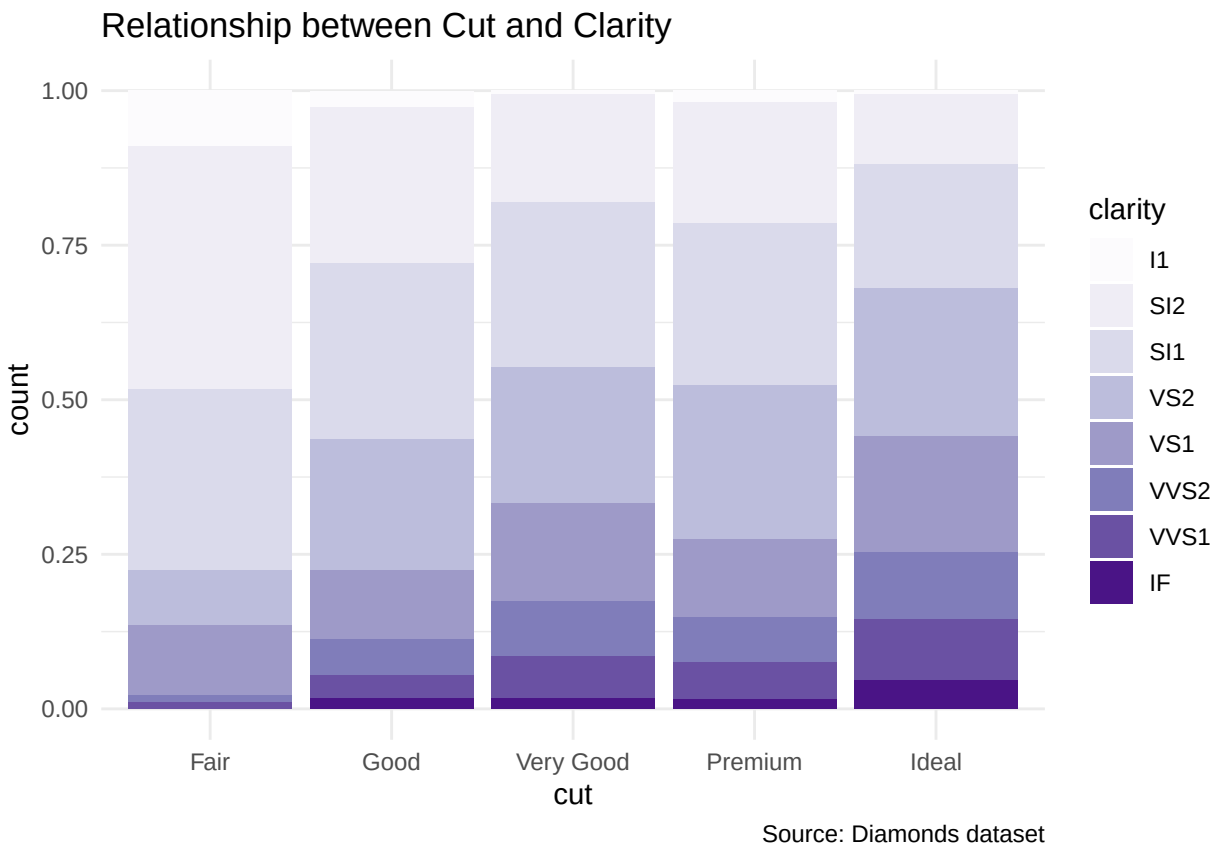


Chart 4: Clarity and Price

```
set.seed(81)
ggplot(sample_n(diamonds,3000), aes(clarity, price , color=clarity)) +
geom_boxplot(outlier.shape = NA) +
labs(title = "Relationship between Diamond clarity and Price",
caption="Source: Diamonds dataset") +
theme_minimal()
```

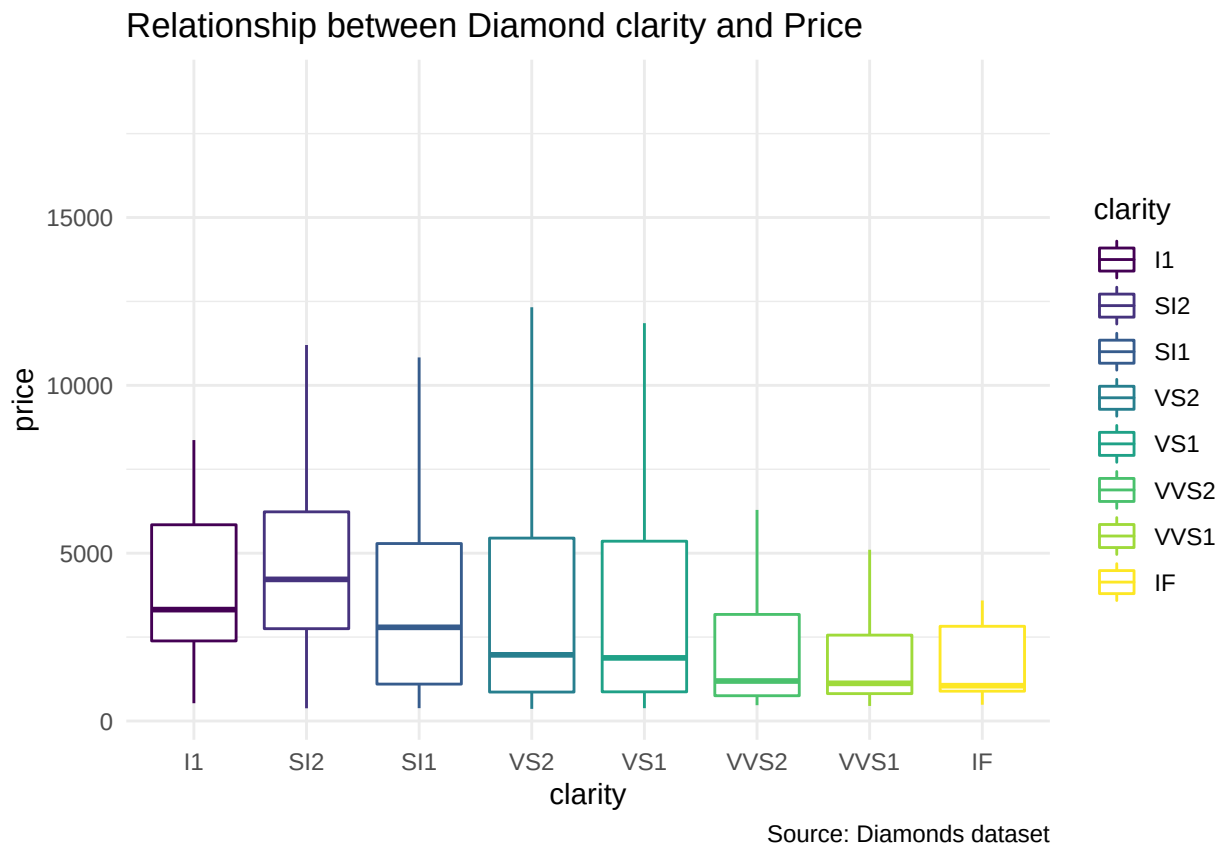
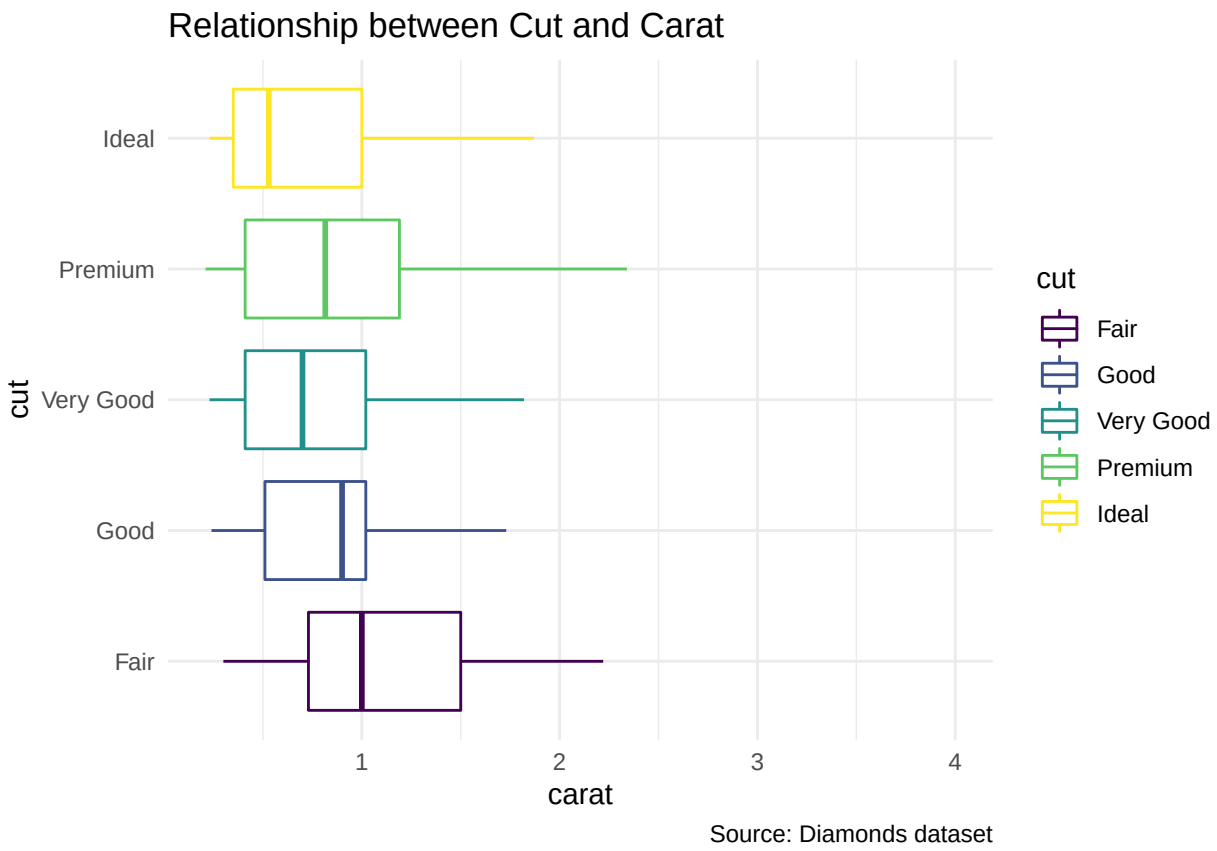


Chart 5: Cut and Carat

```
set.seed(81)
ggplot(sample_n(diamonds, 3000), aes(carat, cut, color = cut )) +
geom_boxplot(outlier.shape = NA) +
labs(title = "Relationship between Cut and Carat",
caption = "Source: Diamonds dataset") +
theme_minimal()
```


Relationship between Cut and Carat
Ideal
Premium
Very Good
Good
Fair
cut
1
2
3
4
carat
cut
Fair
Good
Very Good
Premium
Ideal
Source: Diamonds dataset