

desafio 14

Desafio 14

1. Importe o conjunto de dados `diamonds`, cujo arquivo está disponível no website da disciplina.

```
#reticulate::py_install("scikit-misc")
#install.packages("knitr")
#install.packages("rmarkdown")
```

```
import pandas as pd
from plotnine import *
from plotnine.data import diamonds
from plotnine import ggplot, aes, geom_point, geom_smooth, facet_grid, labs, theme_light, theme

# Verificando as primeiras linhas
print(diamonds.head())
```

	carat	cut	color	clarity	depth	table	price	x	y	z
0	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
1	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
2	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
3	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
4	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75

- 2.

```
from plotnine import ggplot, aes, geom_point, geom_smooth, facet_grid, labs, theme_light, theme

g1 = (
    ggplot(diamonds, aes(x='carat', y='price')) +
    geom_point(alpha=0.3, color="#1f77b4") +
```

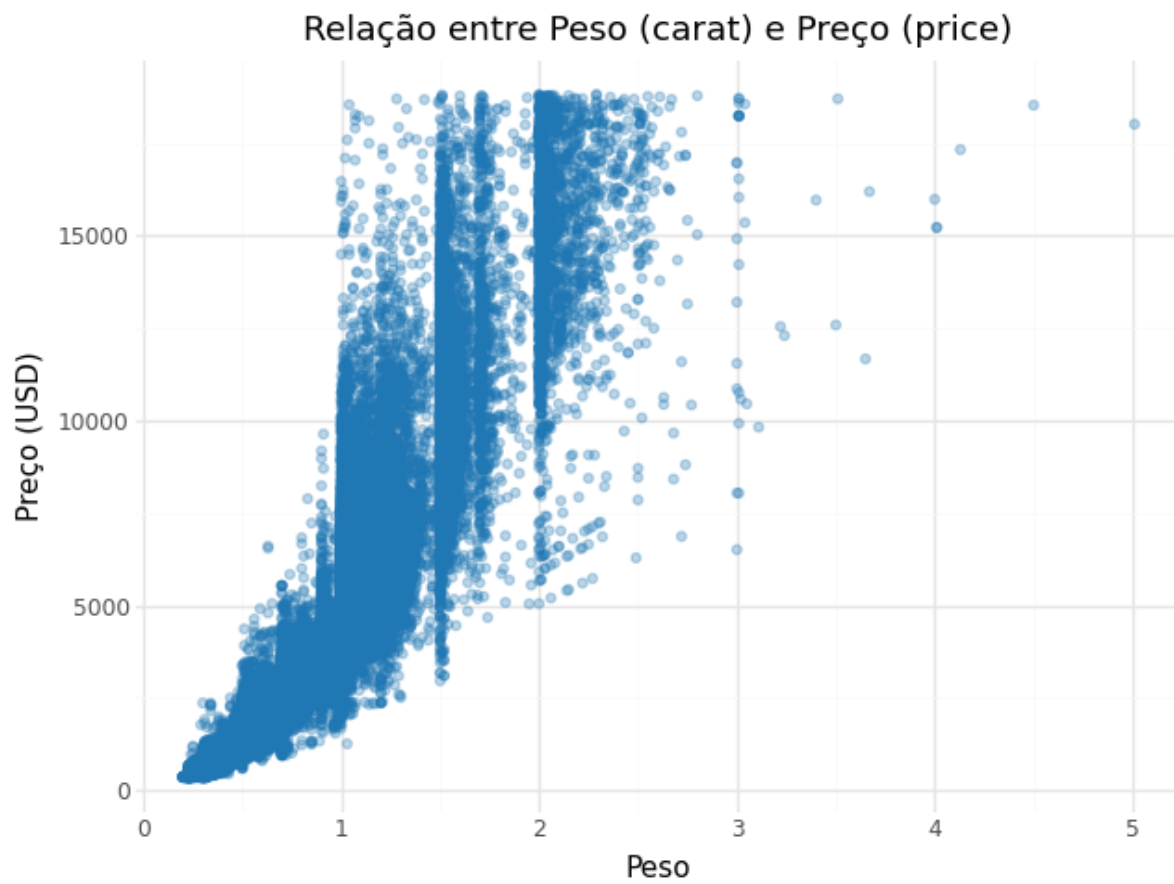
```

labs(
  title='Relação entre Peso (carat) e Preço (price)',
  x='Peso',
  y='Preço (USD)',
) +
theme_minimal()
)

g1.save('grafico_q2.png')

```

\\smb\ra277230\Documentos\.virtualenvs\r-reticulate\Lib\site-packages\plotnine\ggplot.py:630
 \\smb\ra277230\Documentos\.virtualenvs\r-reticulate\Lib\site-packages\plotnine\ggplot.py:631



```

from plotnine import ggplot, aes, geom_point, geom_smooth, facet_grid, labs, theme_light, th
# Gráfico separando por corte
g2 = (
    ggplot(diamonds) +
    aes(x='carat', y='price', color = 'cut') +
    geom_point(alpha=0.3) +
    facet_grid(cols='cut')+
    labs(
        title='Relação entre Peso e Preço por Tipo de Corte',
        x='Peso (carat)',
        y='Preço (USD)',
        color='Corte'
    ) +
    theme_bw()
)

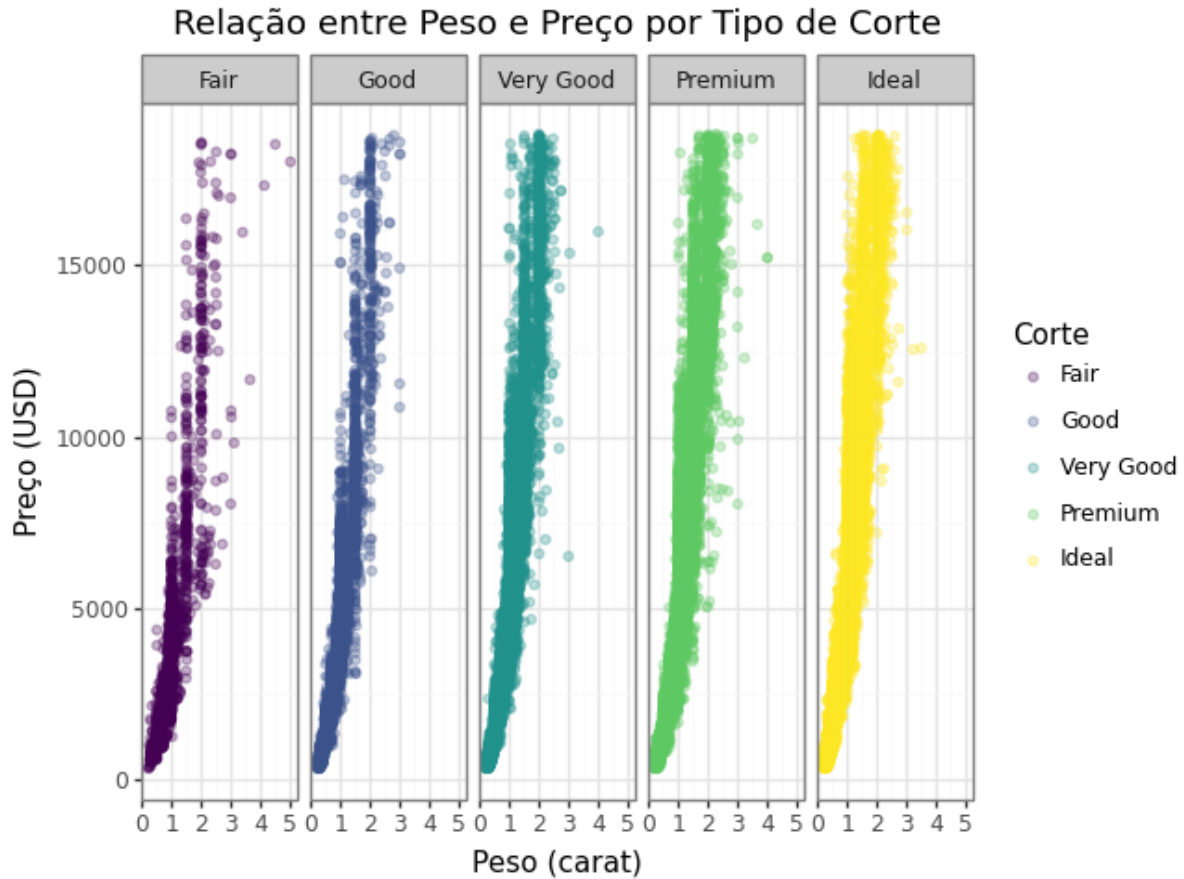
g2.save('grafico_q3.png')

```

```

\\smb\ra277230\Documentos\.virtualenvs\r-reticulate\Lib\site-packages\plotnine\ggplot.py:630
\\smb\ra277230\Documentos\.virtualenvs\r-reticulate\Lib\site-packages\plotnine\ggplot.py:631

```



4

```
# Amostragem para facilitar a visualização (dataset é muito grande)
diamonds_sample = diamonds.sample(5000, random_state=42)

g3 = (
  ggplot(diamonds_sample, aes(x='carat', y='price', color='cut')) +
  geom_point(alpha=0.4, size=1) +
  facet_grid('color ~ clarity') +
  labs(
    title='Relação entre Preço e Peso por Corte, Cor e Clareza',
    x='Peso (carat)',
    y='Preço (USD)'
  ) +
  theme_light() +
  theme(
    legend_position='top',
```

```

    subplots_adjust={'wspace': 0.25}
)
)
print(g3)

```

<ggplot: (640 x 480)>

```
g3.save('grafico_q4.png')
```

\\smb\ra277230\Documentos\.virtualenvs\r-reticulate\Lib\site-packages\plotnine\ggplot.py:630
 \\smb\ra277230\Documentos\.virtualenvs\r-reticulate\Lib\site-packages\plotnine\ggplot.py:631
 \\smb\ra277230\Documentos\.virtualenvs\r-reticulate\Lib\site-packages\plotnine\themes\themeal

