

#Victor de Lima Souza
#199335

```
import pandas as pd
from plotnine import *

diamonds = pd.read_csv("diamonds.csv.gz", compression="gzip")

diamonds.info()
diamonds.head()
```

Executar célula (Ctrl+Enter)
Célula executada desde a última alteração
executada por Victor de Lima Souza
11:16 (há 0 minuto)
executado em 0.006s

```
<class 'pandas.core.frame.DataFrame' >
RangeIndex: 53940 entries, 0 to 53939
Data columns (total 10 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   carat       53940 non-null  float64
1   cut         53940 non-null  object
2   color       53940 non-null  object
3   clarity     53940 non-null  object
4   depth       53940 non-null  float64
5   table       53940 non-null  float64
6   price       53940 non-null  int64
7   x           53940 non-null  float64
8   y           53940 non-null  float64
9   z           53940 non-null  float64
dtypes: float64(6), int64(1), object(3)
memory usage: 4.1+ MB
```

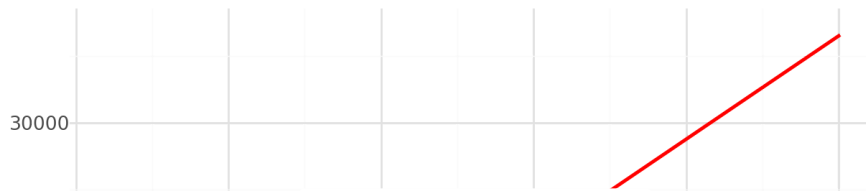
	carat	cut	color	clarity	depth	table	price	x	y	z
0	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
1	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
2	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
3	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
4	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75

Próximas etapas: [Gerar código com diamonds](#) [New interactive sheet](#)

```
(
  ggplot(diamonds, aes(x='carat', y='price')) +
  geom_point(alpha=0.3, color='steelblue') +
  geom_smooth(method='lm', color='red') +
  labs(
    title='Associação entre Peso (carat) e Preço (USD) de Diamantes',
    subtitle='Cada ponto representa um diamante. A linha vermelha indica a tendência linear.',
    x='Peso (carat)',
    y='Preço (USD)'
  ) +
  theme_minimal(base_size=12) +
  theme(
    plot_title=element_text(weight='bold', size=14),
    axis_text_x=element_text(rotation=0)
  )
)
```

Associação entre Peso (carat) e Preço (USD) de Diam

Cada ponto representa um diamante. A linha vermelha indica a tendência



D)

Executar célula (Ctrl+Enter)

Célula executada desde a última alteração

#Para garantir a qualidade do gráfico, a célula foi executada por Victor de Lima Souza em 11:16 (há 0 minuto) em 0.006s

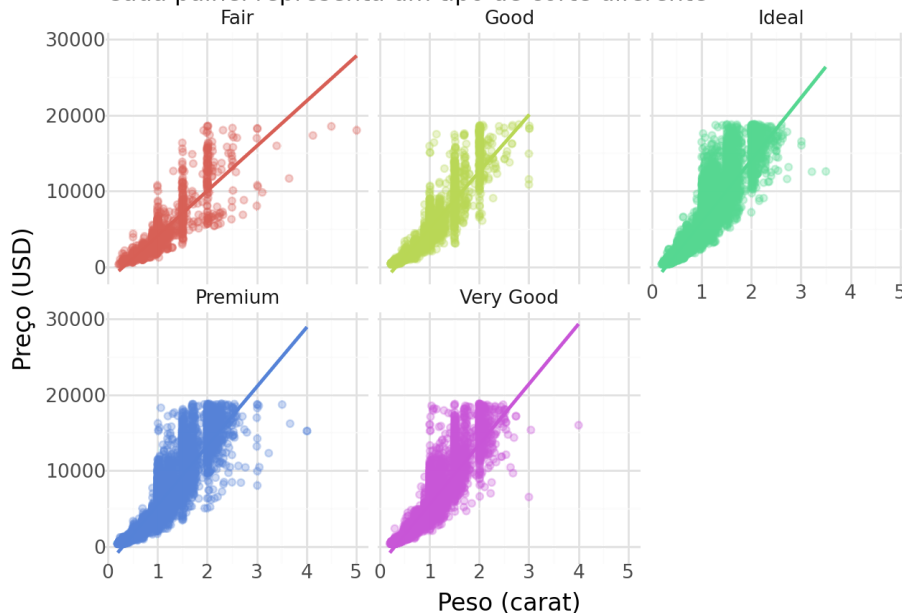
ia para reduzir a sobreposição de pontos, uma linha

#O gráfico revela uma forte associação positiva entre o peso (carat) e o preço dos diamantes. Observa-se que, à medida que o peso aumenta, o preço também tende a aumentar.

```
(
  ggplot(diamonds, aes(x='carat', y='price', color='cut')) +
  geom_point(alpha=0.3) +
  geom_smooth(method='lm', se=False) +
  facet_wrap('~cut') +
  labs(
    title='Relação entre Peso (carat) e Preço (USD) por Tipo de Corte',
    subtitle='Cada painel representa um tipo de corte diferente',
    x='Peso (carat)',
    y='Preço (USD)'
  ) +
  theme_minimal(base_size=12) +
  theme(
    plot_title=element_text(weight='bold', size=14),
    legend_position='none'
  )
)
```

Relação entre Peso (carat) e Preço (USD) por Tipo de Corte

Cada painel representa um tipo de corte diferente



```
diamonds['faixa_carat'] = pd.cut(
    diamonds['carat'],
    bins=[0, 0.5, 1, 2, 3, 5],
    labels=['≤0.5', '0.5-1', '1-2', '2-3', '>3']
)

mean_price = (
    diamonds
    .groupby(['cut', 'faixa_carat'], as_index=False)['price']
    .mean()
    .rename(columns={'price': 'preco_medio'})
)
```

```
mean_price.head()
```

/tmp/ipython-input-3637672398.py:9: FutureWarning: The default of observed=False is deprecated and will be changed

	cut	faixa_carat	preco_medio
--	-----	-------------	-------------

0	Fair	≤0.5	1027.979275
---	------	------	-------------

1	Fair	0.5–1	2802.109661
---	------	-------	-------------

2	Fair	1–2	6127.426966
---	------	-----	-------------

3	Fair	2–3	11697.037
---	------	-----	-----------

4	Fair	>3	14570.777
---	------	----	-----------

Executar célula (Ctrl+Enter)

Célula executada desde a última alteração

executada por Victor de Lima Souza

11:16 (há 0 minuto)

executado em 0.006s

Próximas etapas:

[Gerar código com mean_price](#)[New interactive sheet](#)

#Para aprimorar a visualização, foram utilizadas facetas para comparar a relação entre peso e preço em diferentes

#A análise facetada mostra que, em todos os tipos de corte, a relação entre peso (carat) e preço (price) permanec

```
(
  ggplot(diamonds, aes(x='carat', y='price', color='cut')) +
  geom_point(alpha=0.15, size=1) +
  geom_smooth(method='lm', se=False) +
  facet_grid('color ~ clarity') +
  labs(
    title='Relação entre Preço e Peso do Diamante considerando Corte, Cor e Clareza',
    subtitle='Cada célula combina uma cor e um nível de clareza; as cores indicam o tipo de corte',
    x='Peso (carat)',
    y='Preço (USD)',
    color='Corte'
  ) +
  theme_minimal(base_size=10) +
  theme(
    plot_title=element_text(weight='bold', size=13),
    axis_text_x=element_text(rotation=0),
    figure_size=(12, 8)
  )
)
```

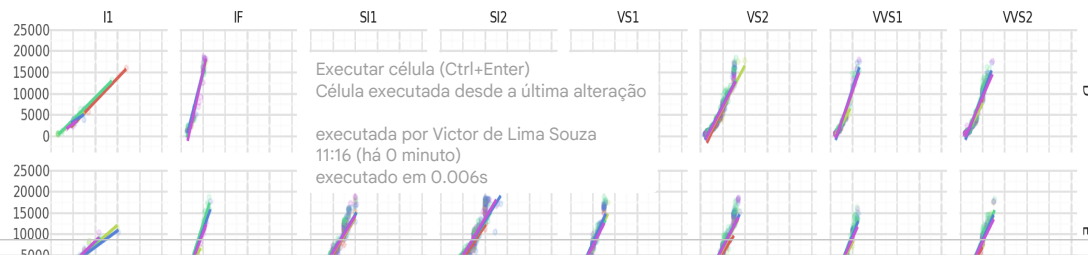
```

/usr/local/lib/python3.12/dist-packages/plotnine/stats/stat_smooth.py:215: PlotnineWarning: Smoothing requires 2 c
/usr/local/lib/python3.12/dist-packages/plotnine/stats/stat_smooth.py:215: PlotnineWarning: Smoothing requires 2 c
/usr/local/lib/python3.12/dist-packages/plotnine/stats/stat_smooth.py:215: PlotnineWarning: Smoothing requires 2 c
/usr/local/lib/python3.12/dist-packages/plotnine/stats/stat_smooth.py:215: PlotnineWarning: Smoothing requires 2 c
/usr/local/lib/python3.12/dist-packages/plotnine/stats/stat_smooth.py:215: PlotnineWarning: Smoothing requires 2 c

```

Relação entre Preço e Peso do Diamante considerando Corte, Cor e Clareza

Cada célula combina uma cor e um nível de clareza; as cores indicam o tipo de corte



#A maior dificuldade encontrada é a visualização do gráfico, pois devido ao seu tamanho e grande quantidade de dados, a relação positiva entre peso e preço continua evidente em todas as combinações de cor, clareza e corte: diamantes de corte Ideal ou Premium apresentam preços consistentemente mais altos, mesmo em pesos semelhantes, e cores próximas de D (melhores) e clarezas como IF e VVS1 também elevam os preços. Já diamantes de cor I/J e clareza SI2/IF mostram uma relação mais dispersa e preços menores. Portanto, o peso é o principal determinante do preço, mas o corte, cor e clareza modulam fortemente o valor, afetando a percepção de qualidade e o valor de mercado.

```

import datetime
import pytz

# Get the current time
now_utc = datetime.datetime.now(datetime.timezone.utc)

# Define the Brasilia timezone
brasilia_timezone = pytz.timezone('America/Sao_Paulo')

# Convert the current time to the Brasilia timezone
now_brasilia = now_utc.astimezone(brasilia_timezone)

# Print the date and time
print("Data e Hora (Brasília):", now_brasilia.strftime('%Y-%m-%d %H:%M:%S %Z'))

```

Data e Hora (Brasília): 2025-10-30 11:16:01 -03-0300