



```
# Instalar o plotnine (se ainda não estiver instalado)
!pip install plotnine
!pip install scikit-misc
# Importar bibliotecas
import pandas as pd
from plotnine import *
from plotnine.data import diamonds
```

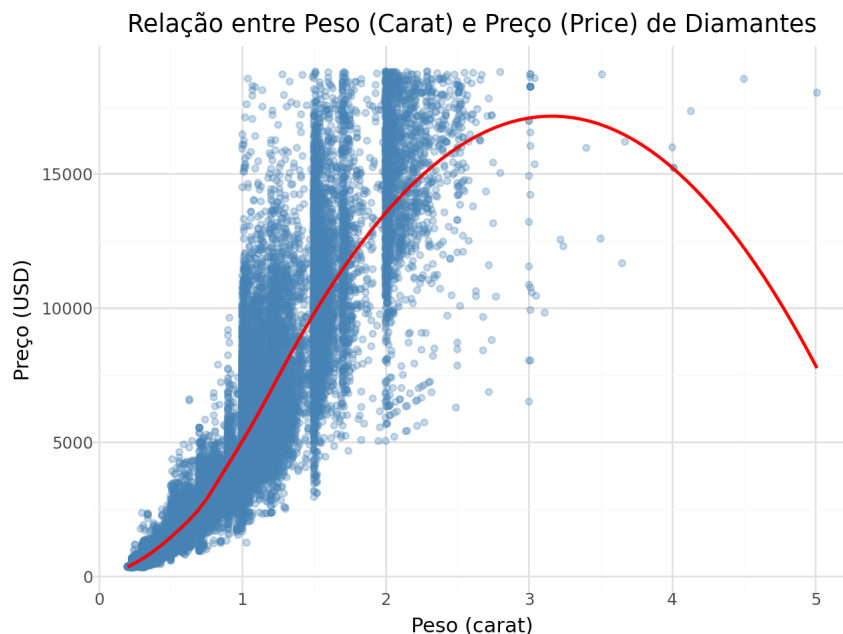
```
# Visualizar o conjunto de dados
diamonds.head()
```

```
Requirement already satisfied: plotnine in /usr/local/lib/python3.12/dist-packages (0.14.5)
Requirement already satisfied: matplotlib>=3.8.0 in /usr/local/lib/python3.12/dist-packages (from plotnine) (3.10.0)
Requirement already satisfied: pandas>=2.2.0 in /usr/local/lib/python3.12/dist-packages (from plotnine) (2.2.2)
Requirement already satisfied: mizani<=0.13.0 in /usr/local/lib/python3.12/dist-packages (from plotnine) (0.13.5)
Requirement already satisfied: numpy>=1.23.5 in /usr/local/lib/python3.12/dist-packages (from plotnine) (2.0.2)
Requirement already satisfied: scipy>=1.8.0 in /usr/local/lib/python3.12/dist-packages (from plotnine) (1.16.2)
Requirement already satisfied: statsmodels>=0.14.0 in /usr/local/lib/python3.12/dist-packages (from plotnine) (0.14.5)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (1.3.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (4.56.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (1.4.7)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (25.0)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.12/dist-packages (from matplotlib>=3.8.0->plotnine) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas>=2.2.0->plotnine) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas>=2.2.0->plotnine) (2025.2)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.12/dist-packages (from statsmodels>=0.14.0->plotnine) (0.5.6)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.7->matplotlib>=3.8.0->plotnine) (1.17.0)
Collecting scikit-misc
  Downloading scikit_misc-0.5.1-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.0 kB)
Requirement already satisfied: numpy>=1.23.5 in /usr/local/lib/python3.12/dist-packages (from scikit-misc) (2.0.2)
Downloading scikit_misc-0.5.1-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (183 kB)
183.6/183.6 kB 6.0 MB/s eta 0:00:00
Installing collected packages: scikit-misc
Successfully installed scikit-misc-0.5.1
```

	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](#)

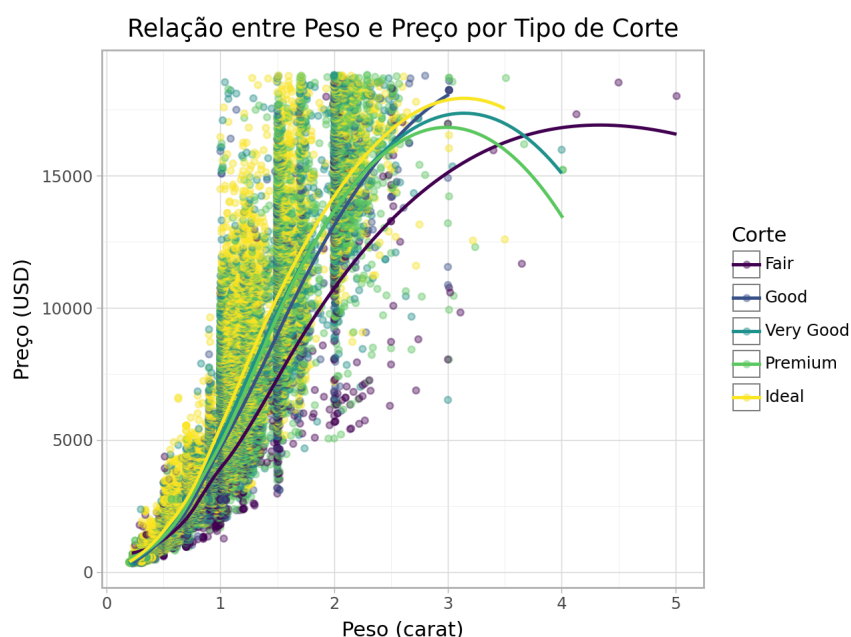
```
# Gráfico de dispersão com suavização (tendência)
(
  ggplot(diamonds, aes(x='carat', y='price')) +
  geom_point(alpha=0.3, color='steelblue') +
  geom_smooth(method='loess', color='red', se=False) +
  labs(
    title='Relação entre Peso (Carat) e Preço (Price) de Diamantes',
    x='Peso (carat)',
    y='Preço (USD)'
  ) +
  theme_minimal()
)
```



Interpretação:

Observa-se uma relação fortemente positiva entre o peso e o preço do diamante. À medida que o peso (carat) aumenta, o preço cresce de forma não linear, sugerindo que diamantes maiores são desproporcionalmente mais caros. Para garantir a qualidade da visualização, utilizei transparência ($\alpha=0.3$) para evitar sobreposição de pontos e adicionei uma curva suavizada (`geom_smooth`) para evidenciar a tendência.

```
(
  ggplot(diamonds, aes(x='carat', y='price', color='cut')) +
  geom_point(alpha=0.4) +
  geom_smooth(method='loess', se=False) +
  labs(
    title='Relação entre Peso e Preço por Tipo de Corte',
    x='Peso (carat)',
    y='Preço (USD)',
    color='Corte'
  ) +
  theme_light()
)
```

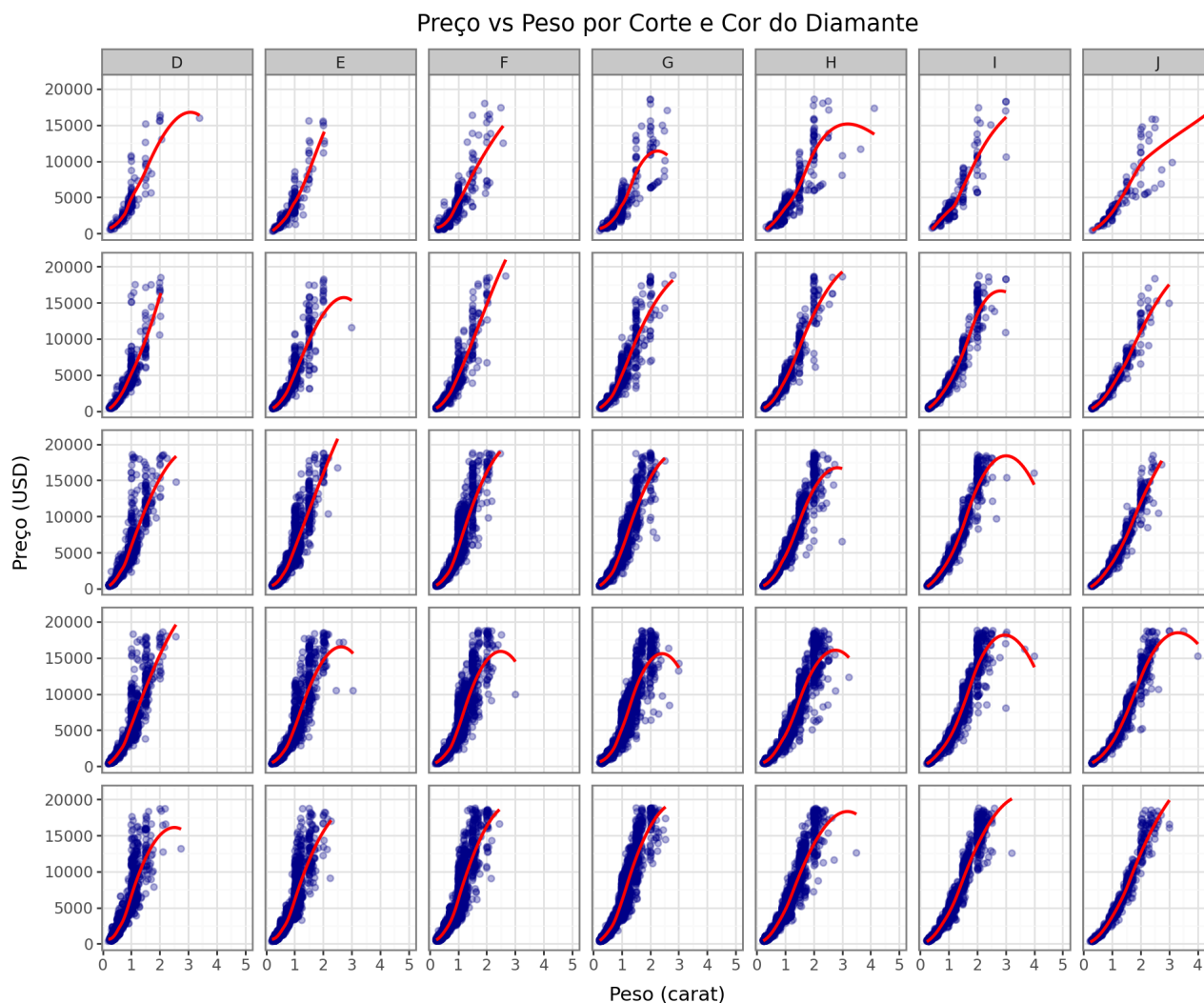


Conclusão:

A relação entre peso e preço é consistente em todos os cortes, mas há diferenças nos níveis de preço: diamantes com cortes melhores (Ideal, Premium) tendem a apresentar preços mais altos para o mesmo peso. O tipo de corte influencia o valor agregado do diamante, não apenas o tamanho físico.

```
(
  ggplot(diamonds, aes(x='carat', y='price')) +
  geom_point(alpha=0.3, color='darkblue') +
  geom_smooth(method='loess', se=False, color='red') +
  facet_grid('cut ~ color') +
  labs(
    title='Preço vs Peso por Corte e Cor do Diamante',
    x='Peso (carat)',
    y='Preço (USD)'
  ) +
  theme_bw() +
  theme(
    figure_size=(10, 8),
    subplots_adjust={'wspace': 0.25, 'hspace': 0.35}
  )
)
```

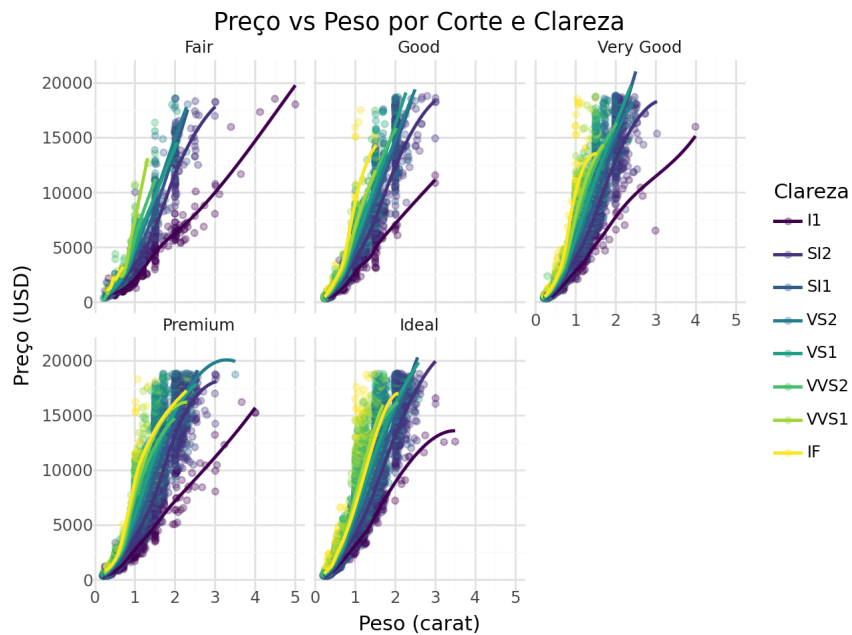
/usr/local/lib/python3.12/dist-packages/plotnine/themes/themeable.py:2438: FutureWarning: You no longer need to use subplots



Conclusão:

Ao combinar corte e cor, nota-se que a relação peso-preço se mantém positiva, mas o nível e a inclinação variam entre categorias. Diamantes com corte Ideal e cor D ou E são mais valorizados. As principais dificuldades envolvem a quantidade de combinações possíveis e o tamanho visual dos gráficos — por isso, foi necessário ajustar o tema e o espaçamento para manter a legibilidade.

```
(
  ggplot(diamonds, aes(x='carat', y='price', color='clarity')) +
  geom_point(alpha=0.3) +
  geom_smooth(method='loess', se=False) +
  facet_wrap('~cut') +
  labs(
    title='Preço vs Peso por Corte e Clareza',
    x='Peso (carat)',
    y='Preço (USD)',
    color='Clareza'
  ) +
  theme_minimal()
)
```



Análise final:

A combinação de corte, cor e clareza explica boa parte da variação no preço. Diamantes de melhor clareza (IF, VVS1) e melhor corte (Ideal, Premium) tendem a custar mais, mesmo com o mesmo peso. A relação é sempre crescente, mas a taxa de crescimento varia conforme a qualidade geral do diamante.