

In [2]:

```
#Brasília - 25-07-2020:

# Introdução:
# Este código analisa Este é o código um banco de dados de venda de veleiros nos EUA.
# O objetivo deste trabalho é mostrar para o velejador quais são os veleiros existente
s,
# relacionando a idade dos barcos por quantidade de anúncios, valor, marca, etc.

# Bons ventos, Cooke.
```

In [5]:

```
# Importando as bibliotecas:

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import matplotlib.pyplot as plt

%matplotlib inline
```

In [6]:

```
# Lendo o Arquivo:
# Os dados foram retirados do site "https://www.boattrader.com/" em 25/07/2020:
data2 = pd.read_csv(r'C:\Users\CookePy\Desktop\datascience\projetos\SailingPrices\query
3\dados_ver9.csv')
```

In [13]:

```
# Observem que temos um banco de dados com 2459 anúncios. Isto representa a totalidade
de anúncios
# existentes no banco de dados da boattrader. Notem a estrutura dos dados. Estamos apen
as começando ...
data2.count()
```

Out[13]:

```
key          2459
Valor        2459
Classe       2459
Ano          2459
Material     2459
Tamanho     2459
Marca        2459
Valor_kUSD   2459
Potencia_cv  2459
dtype: int64
```

In []:

```
# Iniciaremos avaliando os 2459 barcos. Quanto que eles custam e qual o tamanho destes
veleiros?
```

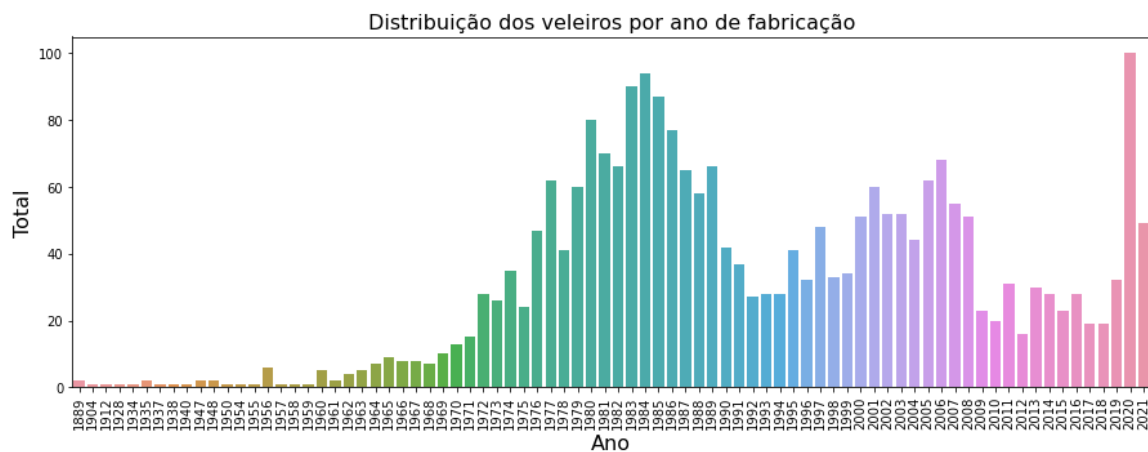
In [22]:

```
dataframe = data2
# Idade média dos veleiros

plt.figure(figsize=(15,5))
sns.countplot(dataframe['Ano'])
plt.xticks(rotation= 90)
plt.ylabel('Total',size=16)
plt.xlabel('Ano', size=16)
plt.title('Distribuição dos veleiros por ano de fabricação', size=16)
```

Out[22]:

Text(0.5, 1.0, 'Distribuição dos veleiros por ano de fabricação')



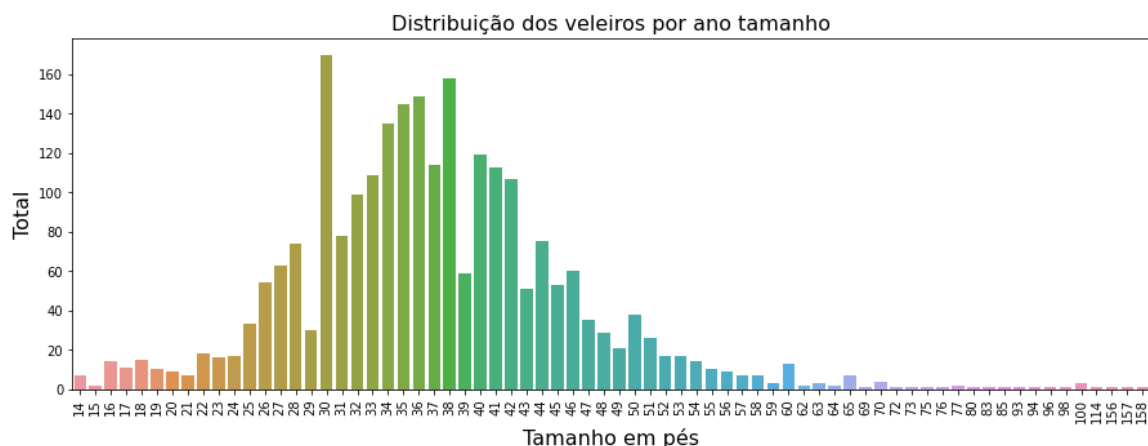
In [23]:

```
dataframe = data2
# Tamanho dos veleiros:

plt.figure(figsize=(15,5))
sns.countplot(dataframe['Tamanho'])
plt.xticks(rotation= 45)
plt.ylabel('Total',size=16)
plt.xlabel('Tamanho em pés', size=16)
plt.title('Distribuição dos veleiros por ano tamanho', size=16)
```

Out[23]:

Text(0.5, 1.0, 'Distribuição dos veleiros por ano tamanho')



In []:

```
# Com os gráficos acima é possível intuir que:
# 1 - A fabricação de veleiros antes de 1970 era quase inexistente.
# 2 - Na década de 80 temos um "boom" na quantidade de veleiros fabricados.
# 3 - Veleiros raramente superam 50 pés, sendo a maior parte dos barcos de 30 a 45 pés
(ou próximo disto..)
```

In [25]:

```
# Eu não vou segurar esta informação. Queremos mesmo saber quanto custa a brincadeira,
não? Vamos plotar os preços:
```

In [38]:

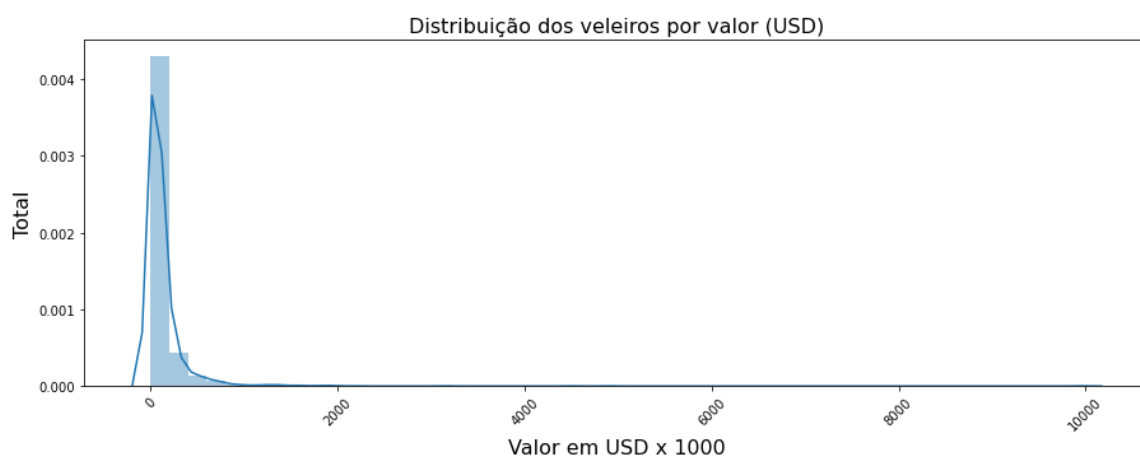
```
dataframe = data2

# Preços dos veleiros:

plt.figure(figsize=(15,5))
sns.distplot(dataframe['Valor_kUSD'])
plt.xticks(rotation= 45)
plt.ylabel('Total',size=16)
plt.xlabel('Valor em USD x 1000', size=16)
plt.title('Distribuição dos veleiros por valor (USD)', size=16)
```

Out[38]:

Text(0.5, 1.0, 'Distribuição dos veleiros por valor (USD)')



In [43]:

```
# O gráfico acima está horrível porque entre as 2459 temos algumas poucas embarcações com valores astronômicos.
# Notem que um veleiro custa menos de 500 mil dólares. Vamos filtrar para entender melhor sobre os preços:
```

In [44]:

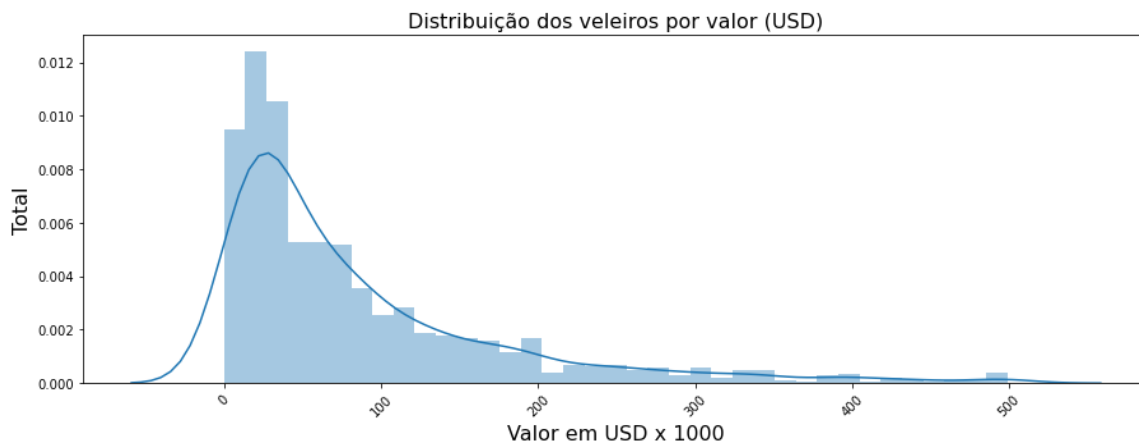
```
query2 = data2.query('Valor_kUSD <500')
dataframe = query2

# Preços dos veleiros:

plt.figure(figsize=(15,5))
sns.distplot(dataframe['Valor_kUSD'])
plt.xticks(rotation= 45)
plt.ylabel('Total',size=16)
plt.xlabel('Valor em USD x 1000', size=16)
plt.title('Distribuição dos veleiros por valor (USD)', size=16)
```

Out[44]:

Text(0.5, 1.0, 'Distribuição dos veleiros por valor (USD)')



In [45]:

```
# Até agora sem muita novidade. Esta primeira parte serviu para fazermos um
# "grande filtro" separando os "out-liers" para que eles não estraguem nossos estudos e
# estatísticos.

# Para as pessoas que entendem menos de estatística: "Não faz sentido comparar um navio
# que custa 10 milhões,
# com veleiros que custam 50 mil dólares.
# Vamos ver onde vai dar..."
```

In []:

```
dataframe = data2
sns.catplot(x="Tamanho", y="Valor", data=data2, jitter = True);
```

In [70]:

```
# ESTE FILTRO ESTA SELECIONANDO BARCOS COM:
# Tamanho de 29 a 50 pés, idade superior a 1970 e valor inferior a 250 mil dólares.
# Notem nos gráficos acima que estamos perdendo poucos dados fazendo este filtro.

df1 = data2.query('29 < Tamanho<50')
df1 = df1.query('Ano>1970')
df1 = df1.query('Valor<250000')
df1 = df1.query('Ano<2020')
```

In [47]:

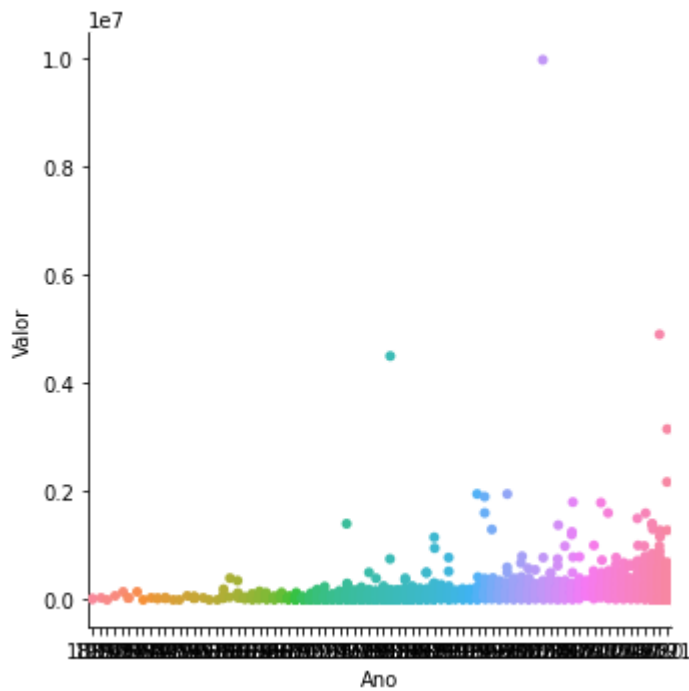
```
# Agora estamos com 1590 anúncios (antes eram 2459). Isto é: anúncios que cumprem os requisitos declarados acima.  
# O número de anúncios foi reduzido em quase mil veleiros, isto se deve a remoção dos veleiros pequenos.  
df1.describe()
```

Out[47]:

	Valor	Ano	Tamanho	Valor_kUSD	Potencia_cv
count	1590.000000	1590.000000	1590.000000	1590.000000	1590.000000
mean	75674.777987	1991.177358	36.907547	75.805031	26.479874
std	56669.914705	11.573934	4.703239	56.647317	26.665006
min	0.000000	1971.000000	30.000000	0.000000	0.000000
25%	29912.500000	1982.000000	33.000000	30.000000	0.000000
50%	59000.000000	1988.000000	36.000000	59.000000	27.000000
75%	109000.000000	2001.000000	40.000000	109.000000	40.000000
max	249900.000000	2019.000000	49.000000	250.000000	475.000000

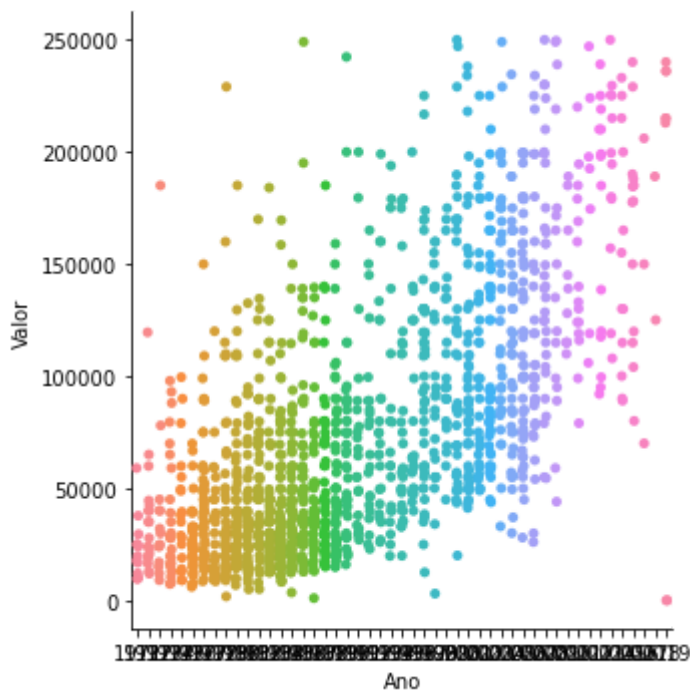
In [89]:

```
# Compare o gráfico abaixo com o proximo, e encontre os OUTLIERS do primeiro.  
# Estou mostrando o mesmo gráfico. No segundo sem outliers. Acho que já pegaram a ideia  
...  
dataframe = data2  
sns.catplot(x="Ano", y="Valor", data=dataframe, jitter = True);
```



In [90]:

```
# É isto. Olha que coerente que ficou. Note que quanto mais novo o barco mais caro ele é. Gênio!
dataframe = df1
sns.catplot(x="Ano", y="Valor", data=dataframe, jitter = True);
```



In []:

In []:

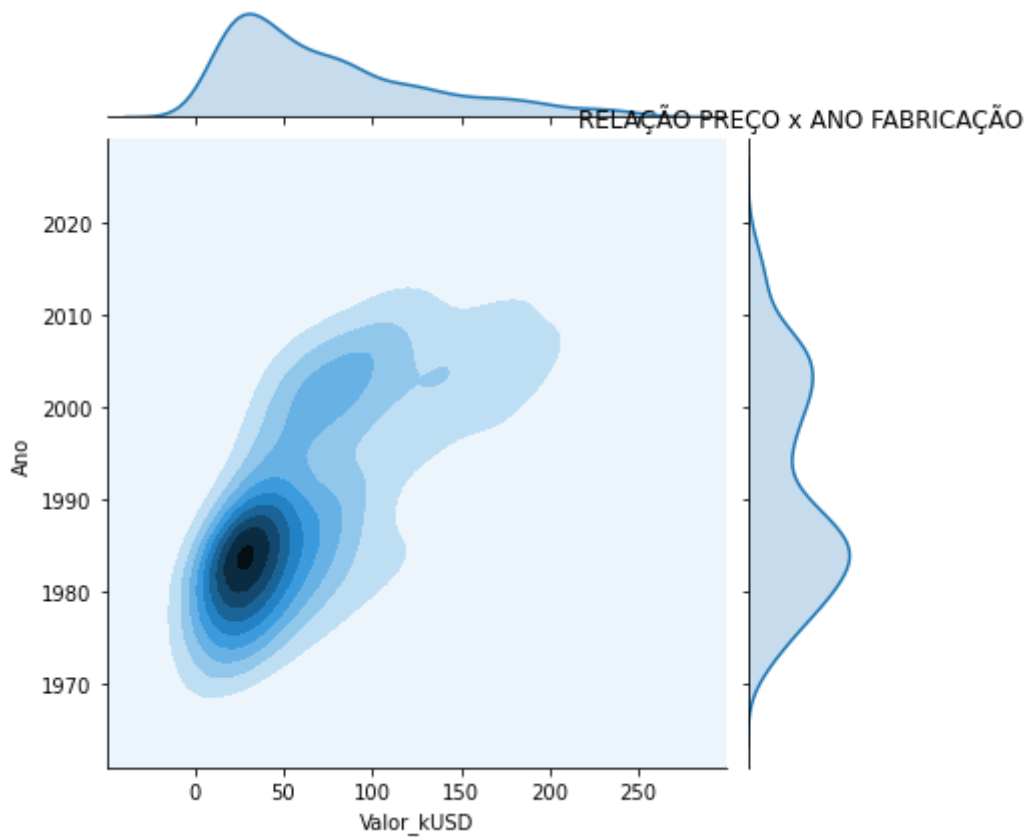
```
#Sem mais demoras, vamos analisar o que interessa:
#Vou te dar um tempo para entender este gráfico. Para as pessoas que não são
#de exatas: A elevação que temos na parte superior e lateral direita mostram a distribu
ição. A figura é a topologia.

# Antes relembrem o que que estamos olhando: Barcos com:# Tamanho de 29 a 50 pés, idade
superior a 1970 e valor inferior a 250 mil dólares.
# Notem nos gráficos acima que estamos perdendo poucos dados fazendo este filtro.
```

In [49]:

```
dataframe = df1
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=dataframe["Valor_kUSD"], y=dataframe["Ano"], kind="kde")
plt.title("RELAÇÃO PREÇO x ANO FABRICAÇÃO")
plt.show()
```

<Figure size 1440x288 with 0 Axes>



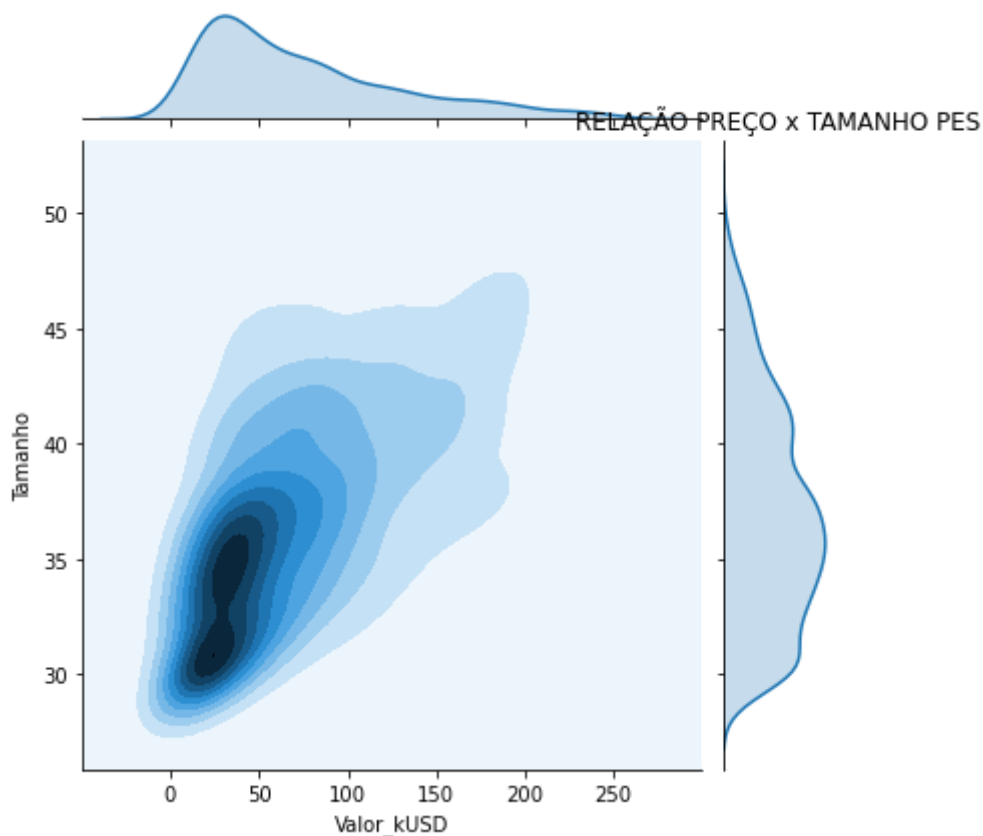
In [50]:

```
# Tvz para o praticante da vela estou mostrando o óbvio, mas pensem no aspirante a velejador,  
# na pessoa sem conhecimento sobre o mercado de nautico. Olhe que genial, podemos intui  
# r muitas coisas!  
# Vamos relacionar as outras variáveis para fechar o cenário:
```

In [52]:

```
dataframe = df1
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=dataframe["Valor_kUSD"], y=dataframe["Tamanho"], kind="kde")
plt.title("RELAÇÃO PREÇO x TAMANHO PES")
plt.show()
```

<Figure size 1440x288 with 0 Axes>



In [55]:

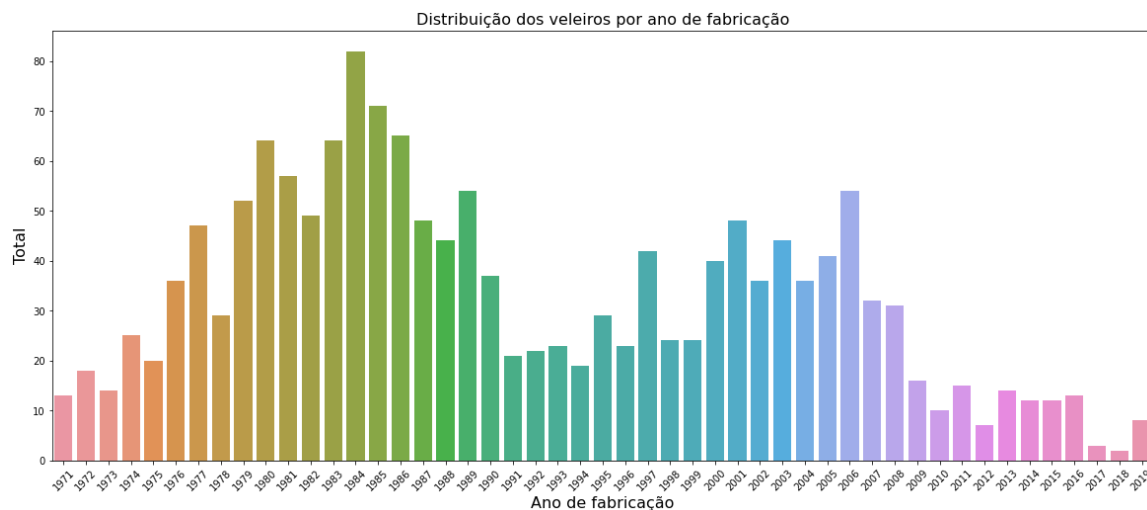
```
# As informações relevantes já estão aí! Espero que tenham gostado.
#Abaixo serão feitas outras análises, um pouco mais do mesmo:
```

In []:

```
# Este gráfico mostra a poluição de veleiros, por tamanho:
```

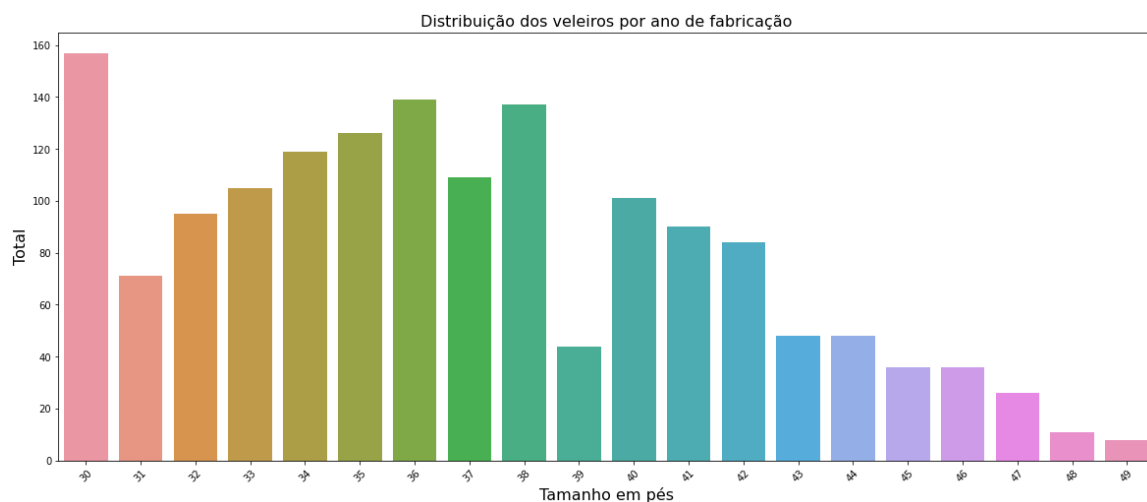

In [93]:

```
dataframe = df1
fig = plt.figure(figsize = (20,8))
sns.countplot(df1["Ano"])
plt.xticks(rotation= 45)
plt.ylabel('Total',size=16)
plt.xlabel('Ano de fabricação', size=16)
plt.title('Distribuição dos veleiros por ano de fabricação', size=16)
plt.show()
```



In [94]:

```
dataframe = df1
fig = plt.figure(figsize = (20,8))
sns.countplot(df1["Tamanho"])
plt.xticks(rotation= 45)
plt.ylabel('Total',size=16)
plt.xlabel('Tamanho em pés', size=16)
plt.title('Distribuição dos veleiros por ano de fabricação', size=16)
plt.show()
```



In [95]:

```
# Agora já temos informações para saber qual é o tamanho e idade do veleiro que "cabe" no bolso.
# Vamos dar uma pirada em marcas, até pq é o que falta para a gnt passar o cartão de crédito..kkk...
```

In [100]:

```
dataframe['Marca'].value_counts()
```

Out[100]:

```
Hunter      213
Catalina    187
Beneteau    140
Pearson     51
C&C         49
...
Alden       1
Homemade    1
Com-Pac     1
Compass     1
Sloop       1
Name: Marca, Length: 263, dtype: int64
```

In []:

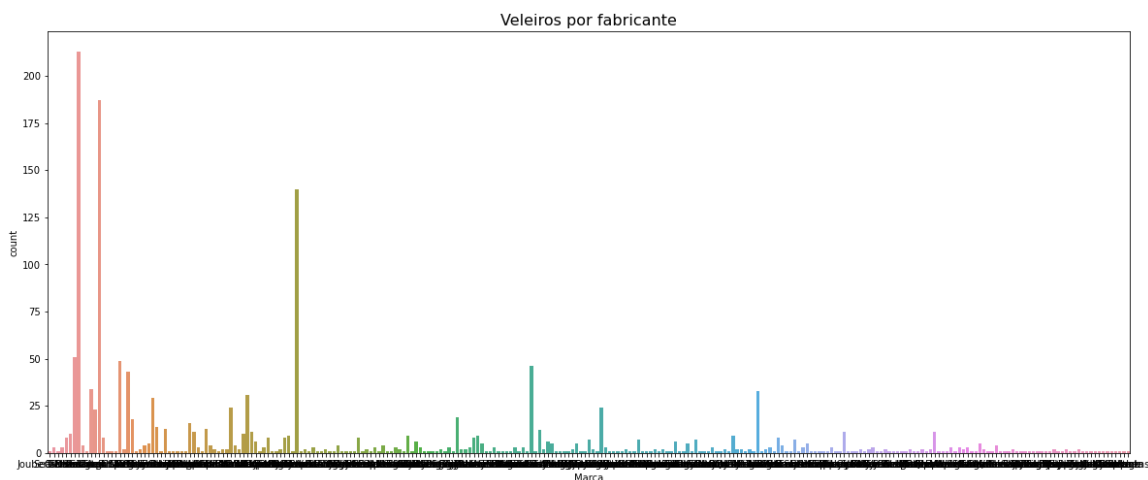
```
# Temos 263 marcas! Mas vejamos no gráfico abaixo, vamos procurar as "grandes"...
```

In [102]:

```
dataframe = df1
fig = plt.figure(figsize = (20,8))
plt.title('Veleiros por fabricante', fontsize = 16)
sns.countplot(dataframe['Marca'])
```

Out[102]:

<matplotlib.axes._subplots.AxesSubplot at 0x1bfc9cb1b08>



In [106]:

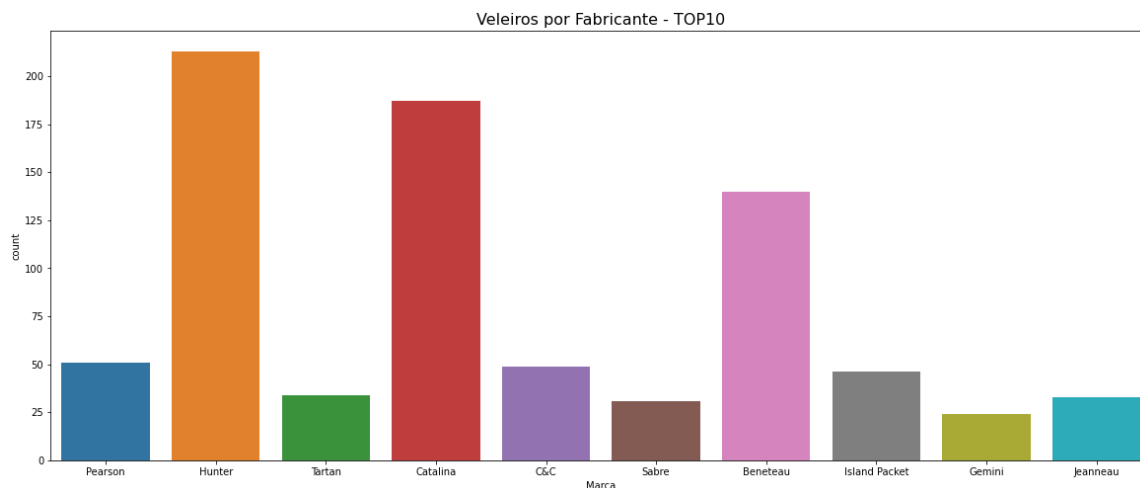
```
dataframe = df1
# Eu vou fazer o filtro para pegar os 10 principais fabricantes. Vamos ver se aparece o Beneteau...
top10fabricantes = dataframe.query('Marca == "Hunter" or Marca == "Catalina" or Marca == "Beneteau" or Marca == "Island Packet" or Marca == "Tartan" or Marca == "C&C" or Marca == "Sabre" or Marca == "Pearson" or Marca == "Jeanneau" or Marca == "Gemini"')
```

In [107]:

```
dataframe = top10fabricantes
fig = plt.figure(figsize = (20,8))
plt.title('Veleiros por Fabricante - TOP10', fontsize = 16)
sns.countplot(dataframe['Marca'])
```

Out[107]:

<matplotlib.axes._subplots.AxesSubplot at 0x1bfc8796948>



In []:

```
# Isto é muito legal. Você já consegue sacar que se puder pegar um Catalina, Hunter ou Beneteau você terá mais peças
# melhor revenda, etc e tal. Podemos fazer umas bricadeiras. Vamos abrir mais o assunto:
```

In [110]:

```
# Recapitulando, este é o nosso filtro: Tamanho de 29 a 50 pés, idade superior a 1970 e valor inferior a 250 mil dólares.
```

In [113]:

```
dataframe = df1
beneteau = dataframe.query('Marca == "Beneteau"')
catalina = dataframe.query('Marca == "Catalina"')
hunter = dataframe.query('Marca == "Hunter"')
```

In [120]:

```
# Abaixo o comparativo das 3 marcas, relacionando o valor com a idade da embarcação.
```

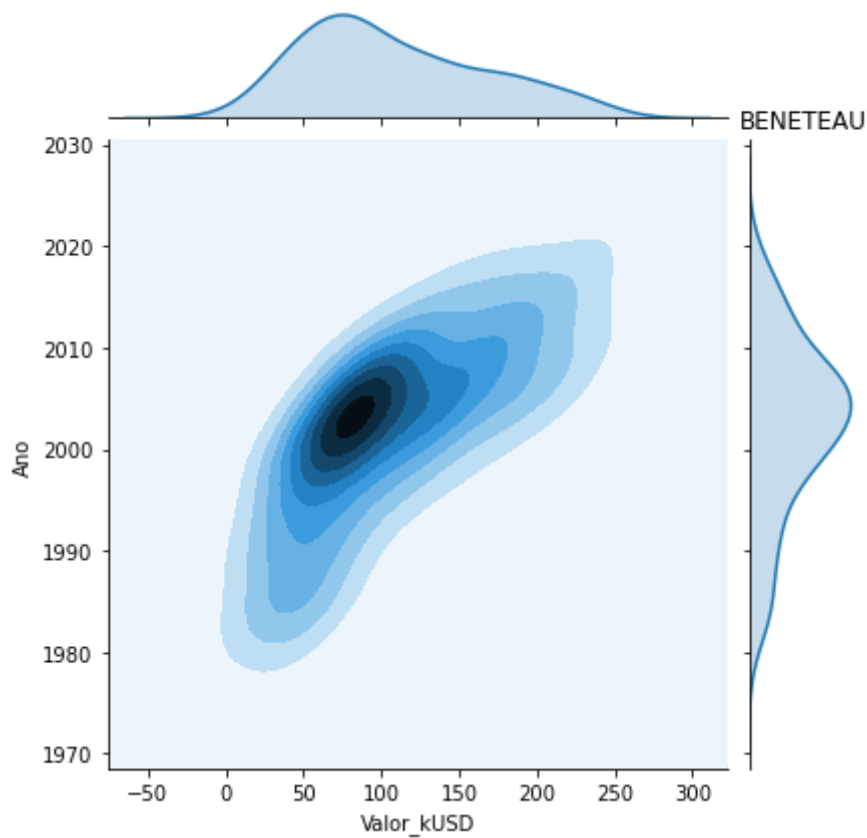
In [121]:

```
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=beneteau["Valor_kUSD"], y=beneteau["Ano"], kind="kde")
plt.title("BENETEAU")
plt.show()

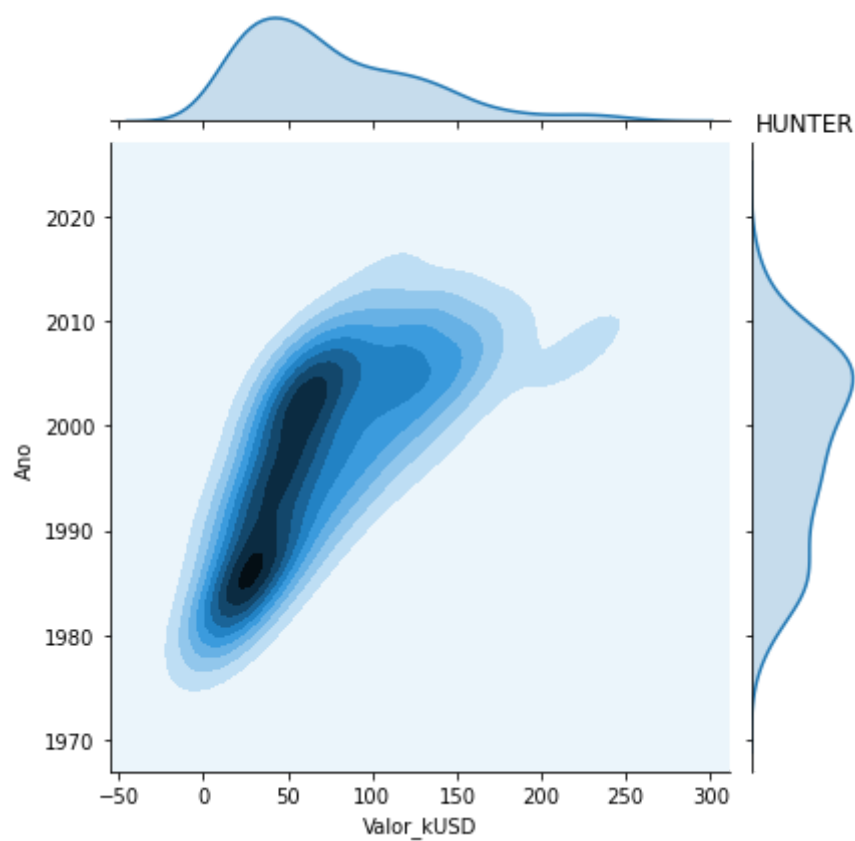
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=hunter["Valor_kUSD"], y=hunter["Ano"], kind="kde")
plt.title("HUNTER")
plt.show()

fig = plt.figure(figsize = (20,4))
sns.jointplot(x=catalina["Valor_kUSD"], y=catalina["Ano"], kind="kde")
plt.title("CATALINA")
plt.show()
```

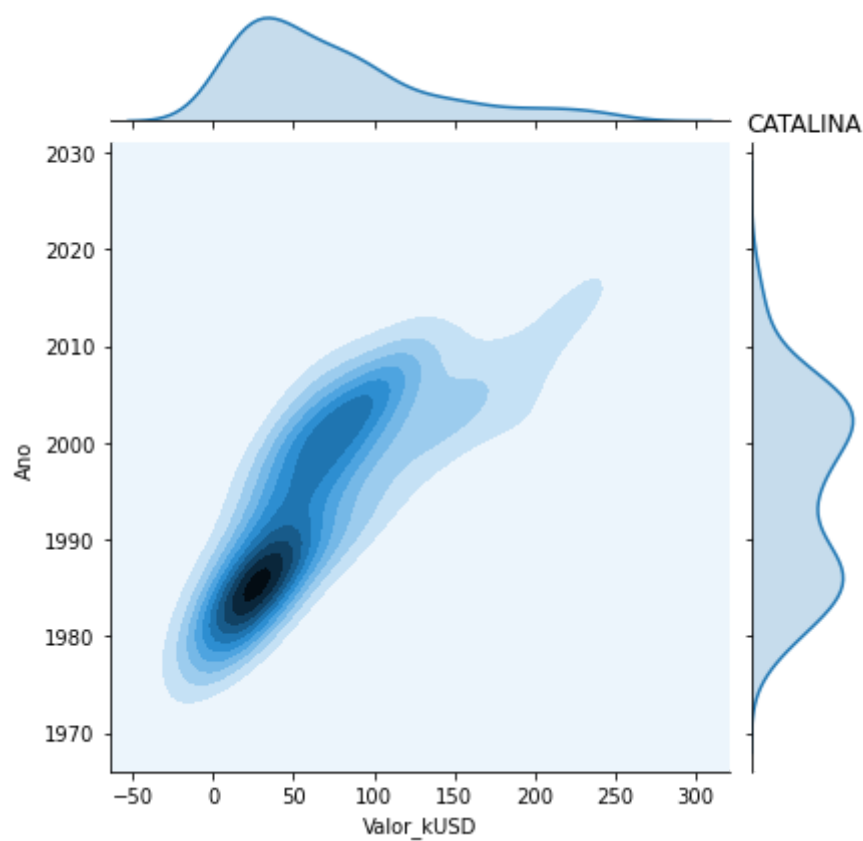
<Figure size 1440x288 with 0 Axes>



<Figure size 1440x288 with 0 Axes>



<Figure size 1440x288 with 0 Axes>



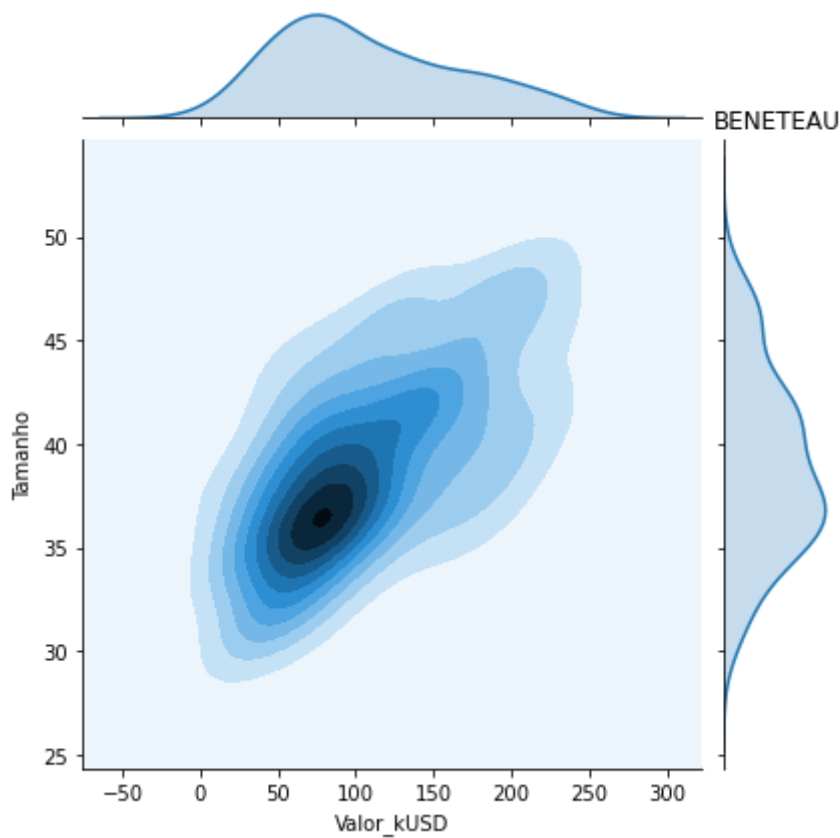
In [122]:

```
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=beneteau["Valor_kUSD"], y=beneteau["Tamanho"], kind="kde")
plt.title("BENETEAU")
plt.show()

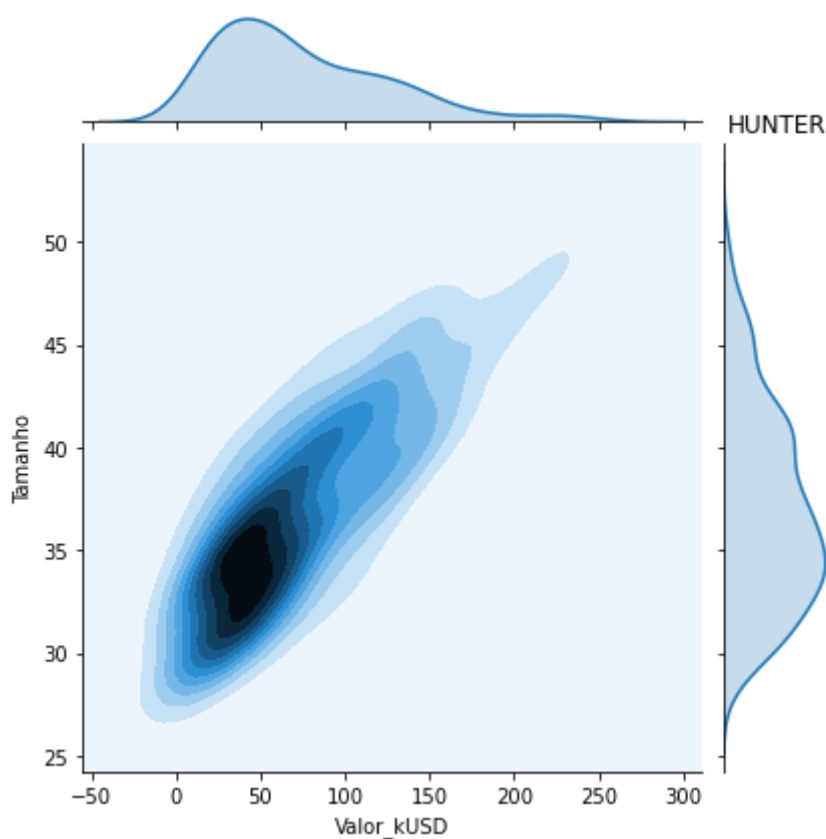
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=hunter["Valor_kUSD"], y=hunter["Tamanho"], kind="kde")
plt.title("HUNTER")
plt.show()

fig = plt.figure(figsize = (20,4))
sns.jointplot(x=catalina["Valor_kUSD"], y=catalina["Tamanho"], kind="kde")
plt.title("CATALINA")
plt.show()
```

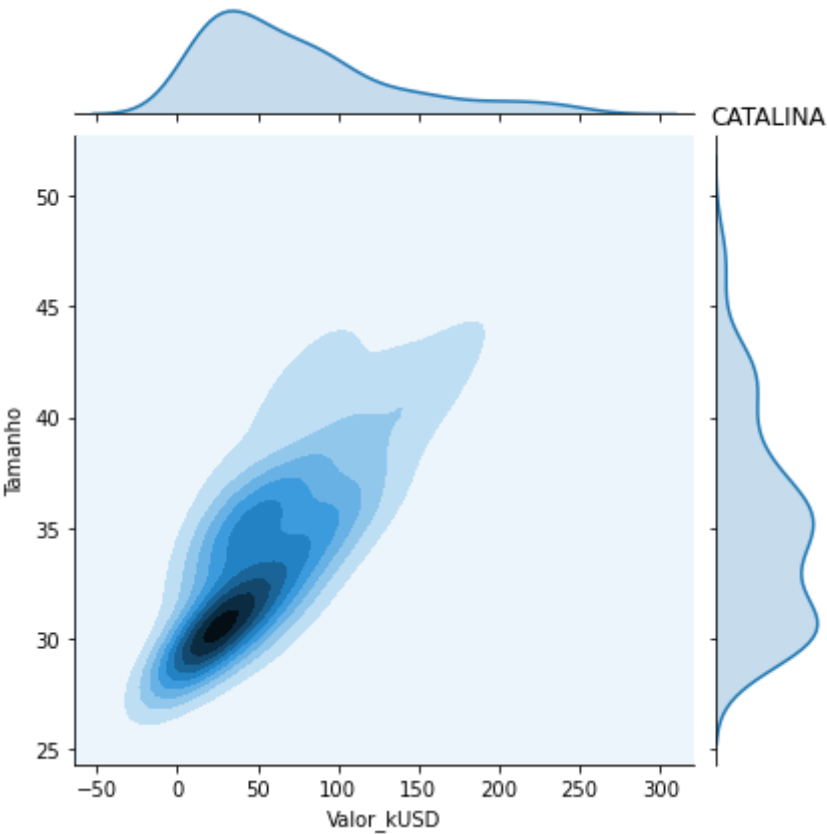

<Figure size 1440x288 with 0 Axes>



<Figure size 1440x288 with 0 Axes>



<Figure size 1440x288 with 0 Axes>



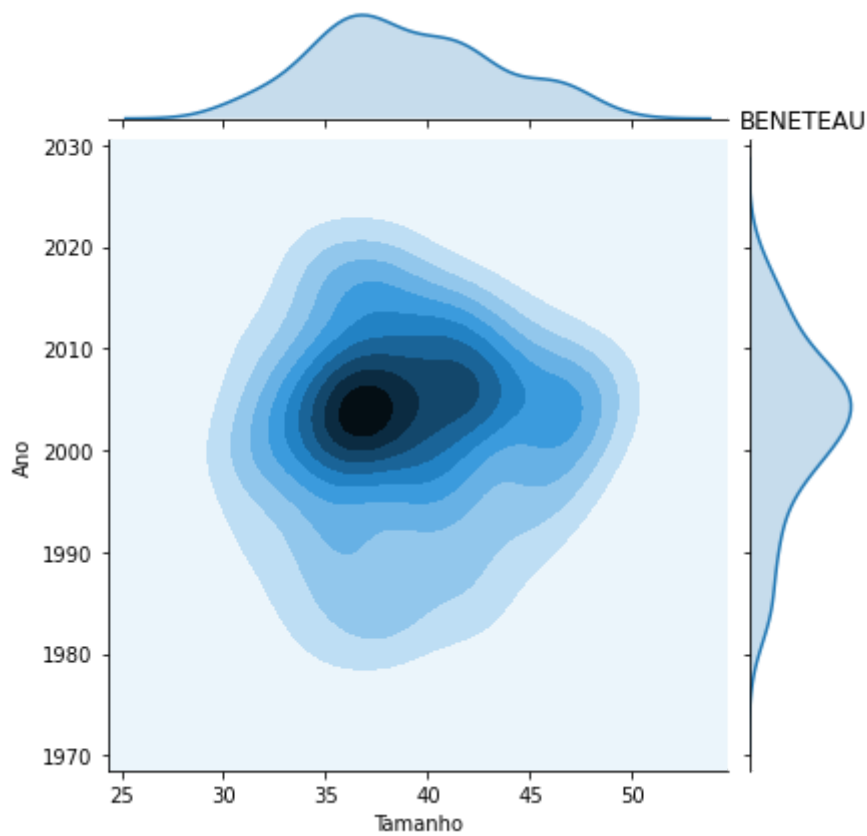
In [123]:

```
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=beneteau["Tamanho"], y=beneteau["Ano"], kind="kde")
plt.title("BENETEAU")
plt.show()

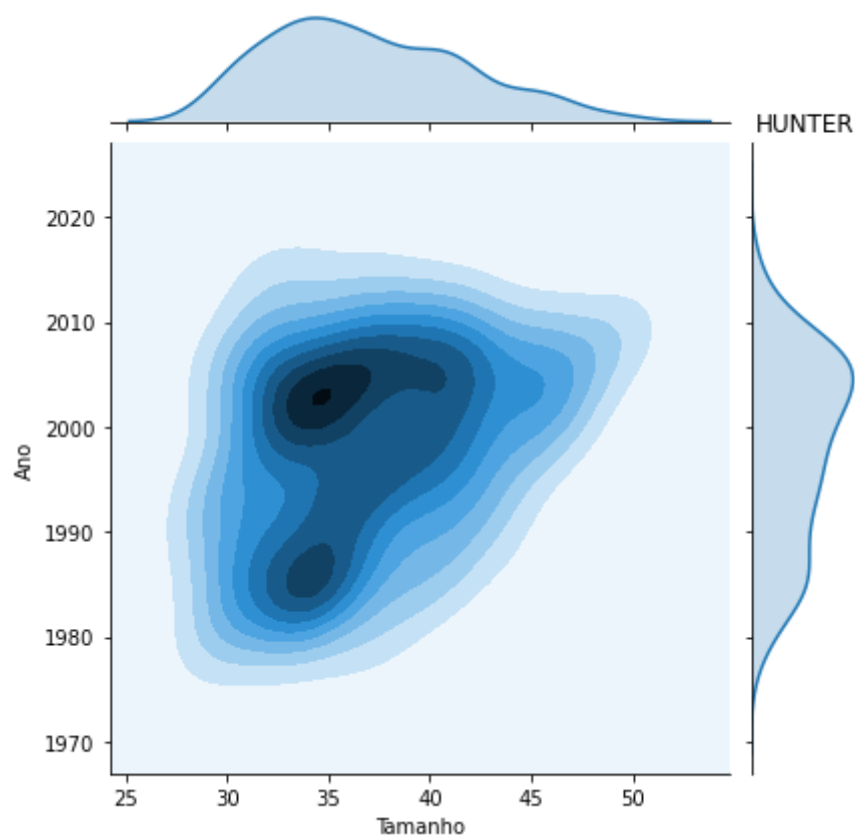
fig = plt.figure(figsize = (20,4))
sns.jointplot(x=hunter["Tamanho"], y=hunter["Ano"], kind="kde")
plt.title("HUNTER")
plt.show()

fig = plt.figure(figsize = (20,4))
sns.jointplot(x=catalina["Tamanho"], y=catalina["Ano"], kind="kde")
plt.title("CATALINA")
plt.show()
```

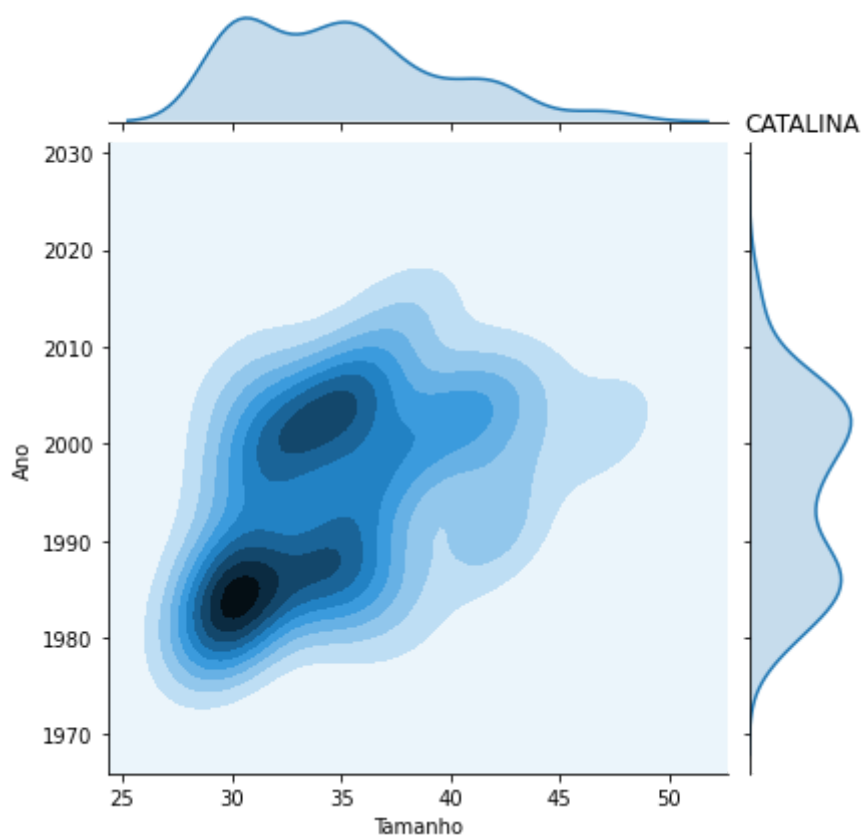
<Figure size 1440x288 with 0 Axes>



<Figure size 1440x288 with 0 Axes>



<Figure size 1440x288 with 0 Axes>



In []:

```
lista_marcas_cooke = data3['Marca'].value_counts()
```

In []:

```
# É isto aí, pessoal.  
#Da para brincar mais, e na próxima versão eu vou apresentar alguns  
#dados interessantes sobre outras características das embarcações.  
#vou fechar com um exemplo. Obrigado pelo seu tempo, bons ventos!! Cooke!
```

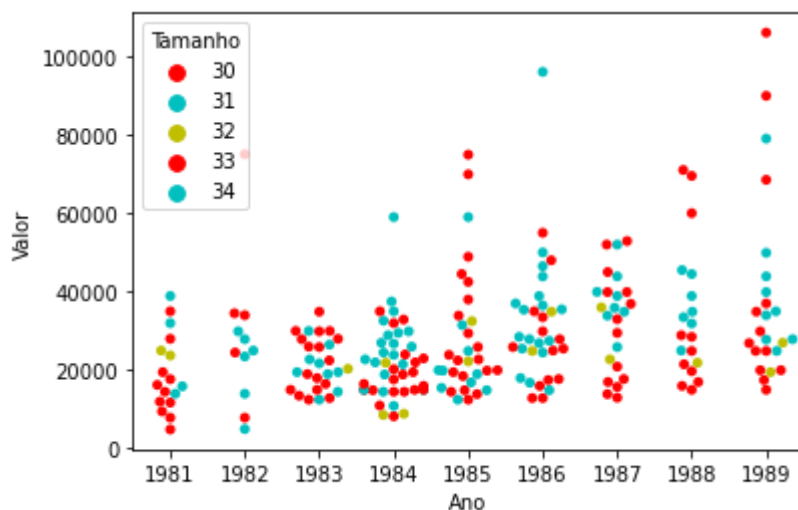
In [131]:

```
# Veleiros da década dourada da vela, no tamanho "Possível", por volta dos 30 pés!
```

```
decada_dourada = dataframe.query('1980 < Ano < 1990')  
boat_size = decada_dourada.query('29 < Tamanho < 35')  
sns.swarmplot(x="Ano", y="Valor", hue="Tamanho",  
              palette=["r", "c", "y"], data=boat_size)
```

Out[131]:

<matplotlib.axes._subplots.AxesSubplot at 0x1bfd14e1e88>



In []: