

Pergunta 2. Qual bandeira de postos de combustíveis apresenta os menores preços?

<https://altair-viz.github.io/> (<https://altair-viz.github.io/>)

Carregar dados

In [1]:

```
import pandas as pd
import altair as alt
```

In [2]:

```
from funcoes_apoio import preparar_dataframe
```

In [3]:

```
df = preparar_dataframe(agrupar=True)
df
```

Out[3]:

	regiao	uf	produto	bandeira	ano	mes	ams	venda	compra	margem
0	S	PR	GASOLINA	BRANCA	2015	1	201501	3.007487	2.619785	0.388
72	S	PR	GASOLINA	BRANCA	2015	2	201502	3.279594	2.795547	0.484
144	S	PR	GASOLINA	BRANCA	2015	3	201503	3.310976	2.885765	0.425
216	S	PR	GASOLINA	BRANCA	2015	4	201504	3.275583	2.875055	0.401
288	S	PR	GASOLINA	BRANCA	2015	5	201505	3.279458	2.867840	0.412
...
6900840	SE	SP	ETANOL	RAIZEN	2020	7	202007	2.590275	2.201664	0.389
6900912	SE	SP	ETANOL	RAIZEN	2020	8	202008	2.608820	2.247416	0.361
6901055	SE	SP	ETANOL	RAIZEN	2020	10	202010	2.893640	2.493020	0.401
6901127	SE	SP	ETANOL	RAIZEN	2020	11	202011	2.984247	2.601998	0.382
6901199	SE	SP	ETANOL	RAIZEN	2020	12	202012	3.052666	2.696621	0.356

18059 rows × 11 columns

Agrupar dados

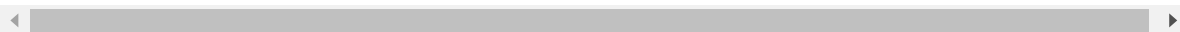
In [11]:

```
df2 = df.\
    groupby([
        pd.Grouper(key='regiao'),
        pd.Grouper(key='uf'),
        pd.Grouper(key='produto'),
        pd.Grouper(key='ano'),
        pd.Grouper(key='mes'),
    ]).mean().reset_index().dropna(subset=['venda'])
df2
```

Out[11]:

	regiao	uf	produto	ano	mes	ams	venda	compra	margem	pmargem
0	S	PR	GASOLINA	2015	1	201501.0	3.014511	2.634961	0.3798	14.404000
1	S	PR	GASOLINA	2015	2	201502.0	3.289405	2.818746	0.4706	16.698000
2	S	PR	GASOLINA	2015	3	201503.0	3.328853	2.908722	0.4202	14.442000
3	S	PR	GASOLINA	2015	4	201504.0	3.295943	2.904184	0.3920	13.490000
4	S	PR	GASOLINA	2015	5	201505.0	3.297062	2.893235	0.4040	13.958000
...
19434	SE	SP	ETANOL	2020	7	202007.0	2.587264	2.200899	0.3864	17.543999
19435	SE	SP	ETANOL	2020	8	202008.0	2.603317	2.242355	0.3608	16.090000
19437	SE	SP	ETANOL	2020	10	202010.0	2.872609	2.486333	0.3864	15.532000
19438	SE	SP	ETANOL	2020	11	202011.0	2.972598	2.589883	0.3828	14.776000
19439	SE	SP	ETANOL	2020	12	202012.0	3.030175	2.679195	0.3510	13.092000

3819 rows × 10 columns



In [23]:

```
df3 = df.\n    groupby([\n        pd.Grouper(key='regiao'),\n        #pd.Grouper(key='uf'),\n        pd.Grouper(key='produto'),\n        pd.Grouper(key='bandeira'),\n        pd.Grouper(key='ano'),\n        pd.Grouper(key='mes'),\n    ]).mean().reset_index().dropna(subset=['venda'])\ndf3['ams'] = df3['ano'].astype('int') * 100 + df3['mes'].astype('int')\ndf3['data'] = pd.to_datetime(df3['ams'], format='%Y%m')\ndf3
```

Out[23]:

	regiao	produto	bandeira	ano	mes	ams	venda	compra	margem	pmarg
0	S	GASOLINA	BRANCA	2015	1	201501	3.004279	2.606254	0.398000	15.2760
1	S	GASOLINA	BRANCA	2015	2	201502	3.294508	2.795362	0.499000	17.8560
2	S	GASOLINA	BRANCA	2015	3	201503	3.298532	2.877332	0.421000	14.6360
3	S	GASOLINA	BRANCA	2015	4	201504	3.246368	2.854174	0.392333	13.7400
4	S	GASOLINA	BRANCA	2015	5	201505	3.241515	2.842692	0.398667	14.0300
...
3594	SE	ETANOL	RAIZEN	2020	7	202007	3.173419	2.724994	0.448500	16.6800
3595	SE	ETANOL	RAIZEN	2020	8	202008	3.193010	2.758320	0.434500	15.9270
3597	SE	ETANOL	RAIZEN	2020	10	202010	3.318358	2.911033	0.407500	14.1920
3598	SE	ETANOL	RAIZEN	2020	11	202011	3.435342	3.017002	0.418250	13.9270
3599	SE	ETANOL	RAIZEN	2020	12	202012	3.513423	3.104527	0.408750	13.1670

3550 rows × 11 columns

Seletores

In [12]:

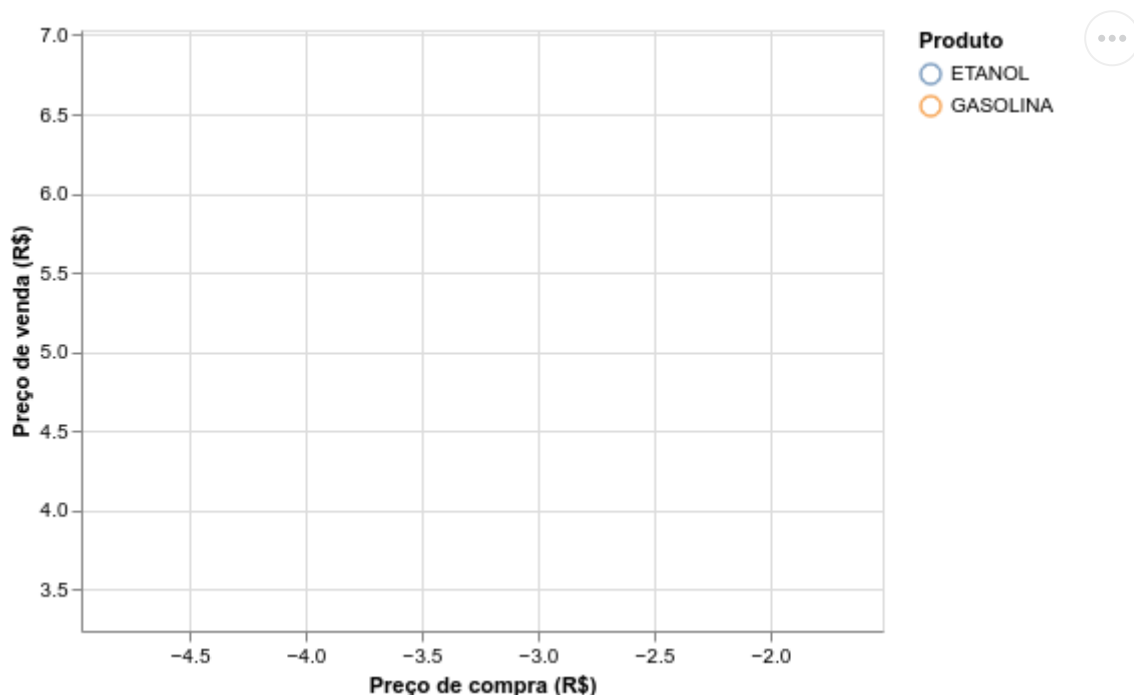
```
# criar seletor de ano
menor, maior = df['ano'].min(), df['ano'].max()
seletor_ano = alt.selection_single(
    name='Select', fields=['ano'], init={'ano': menor},
    bind=alt.binding_range(min=menor, max=maior, step=1)
)
```

Preço de compra vs Preço de venda

In [13]:

```
alt.Chart(df2).mark_point().encode(
    alt.X('compra', title='Preço de compra (R$)'),
    alt.Y('venda', title='Preço de venda (R$)'),
    alt.Color('produto', title='Produto'),
    tooltip = [
        alt.Tooltip('produto:N', title='Produto'),
        #alt.Tooltip('municipio:N', title='Município'),
        alt.Tooltip('uf:N', title='UF'),
        alt.Tooltip('compra:Q', title="Preço de compra (R$)", format=".3f"),
        alt.Tooltip('venda:Q', title="Preço de venda (R$)", format=".3f"),
    ]
).interactive()
```

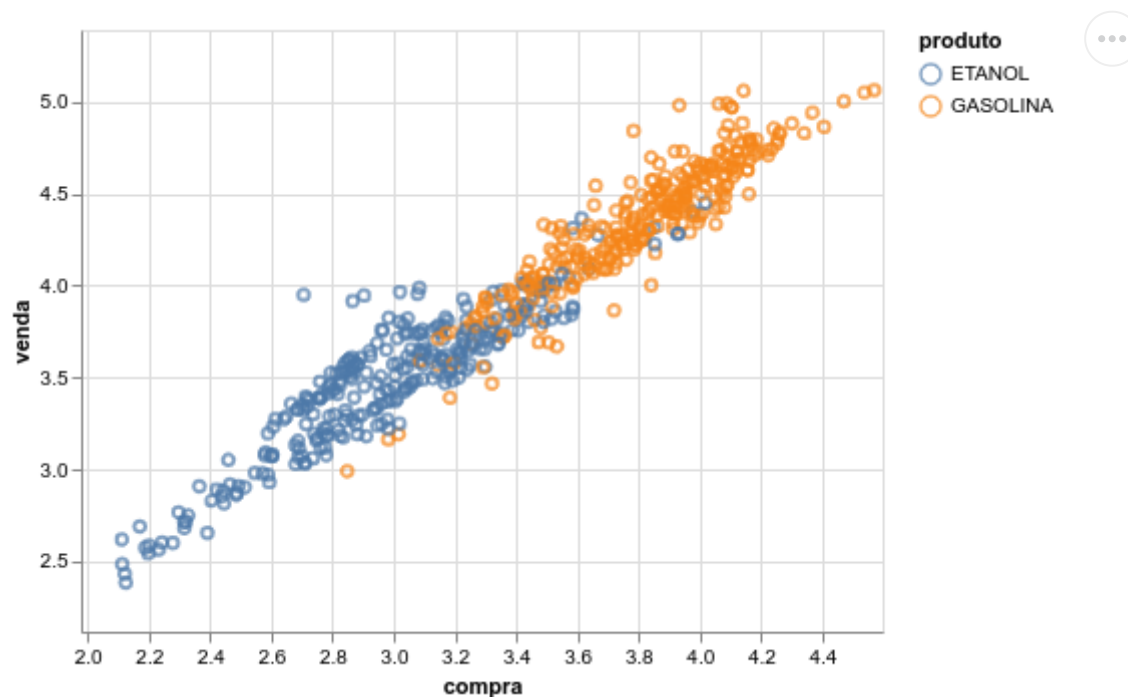
Out[13]:



In [14]:

```
alt.Chart(df2).mark_point().encode(  
    x='compra',  
    y='venda',  
    color='produto',  
    tooltip=['produto', 'uf', 'compra', 'venda', 'margem', 'pmargem']  
)  
.add_selection(seletor_ano).transform_filter(seletor_ano).interactive()
```

Out[14]:



Select_ano



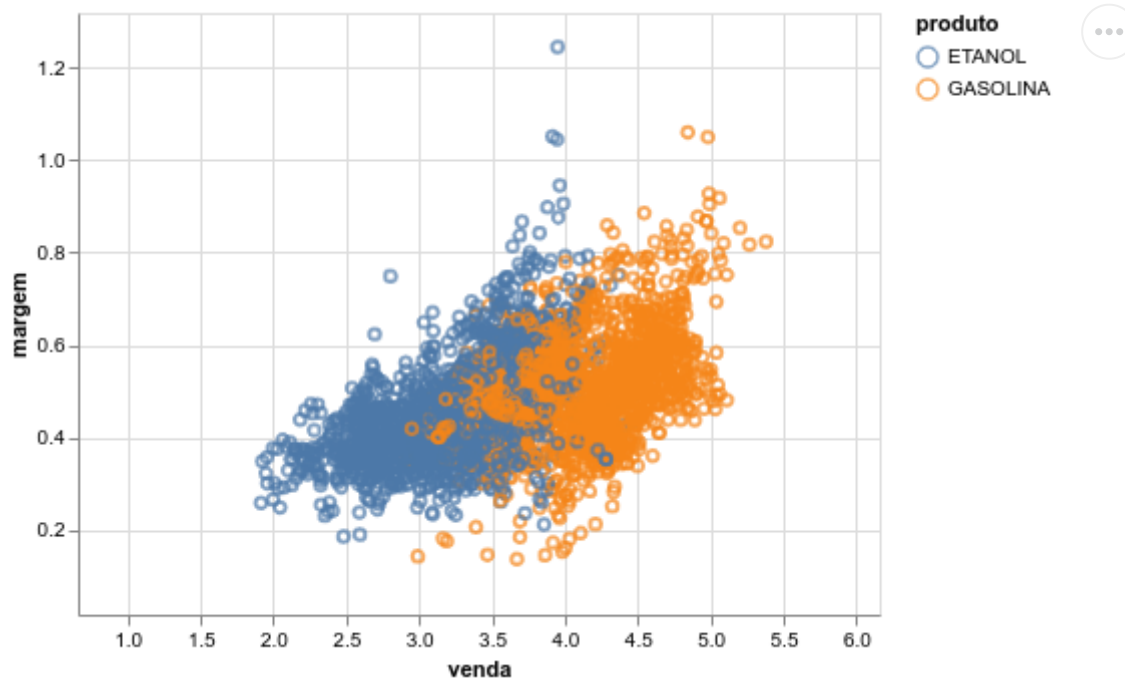
2020

Preço de venda vs Lucro

In [15]:

```
alt.Chart(df2).mark_point().encode(  
    x='venda',  
    y='margem',  
    color='produto',  
    tooltip=['uf', 'ano', 'compra', 'venda', 'margem', 'pmargem']  
)
```

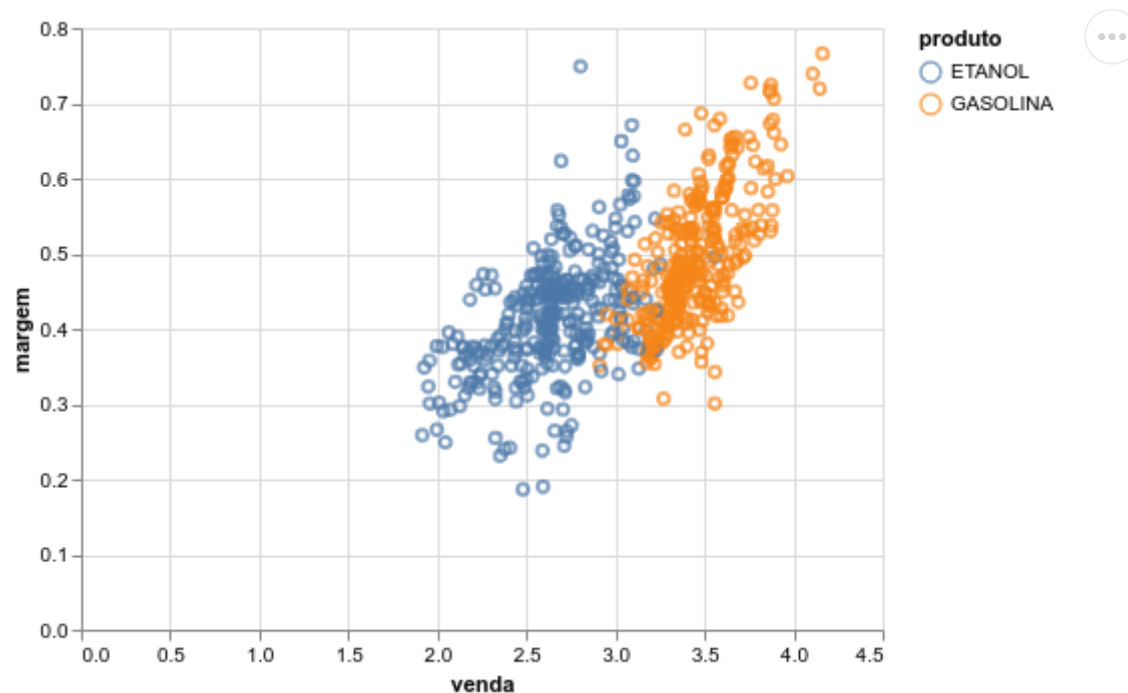
Out[15]:



In [16]:

```
alt.Chart(df2).mark_point().encode(  
    x='venda',  
    y='margem',  
    color='produto',  
    tooltip=['produto', 'uf', 'compra', 'venda', 'margem', 'pmargem']  
)  
.add_selection(seletor_ano).transform_filter(seletor_ano).interactive()
```

Out[16]:



Select_ano



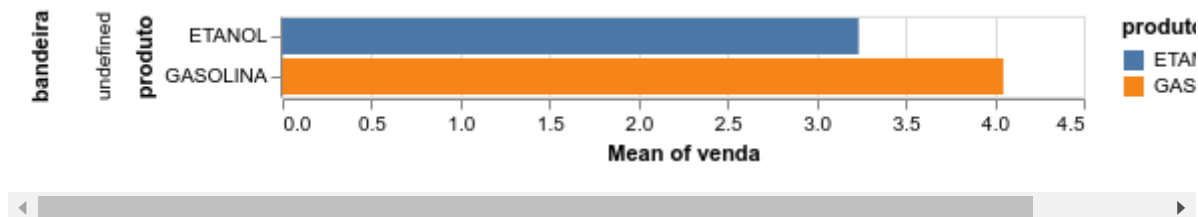
2015

...

In [17]:

```
alt.Chart(df2).mark_bar().encode(
    x='mean(venda)',
    y='produto:N',
    color='produto:N',
    row='bandeira:N'
)
```

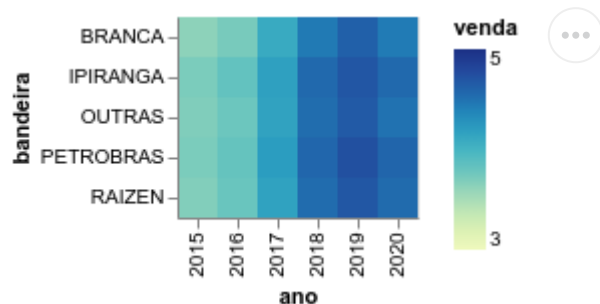
Out[17]:



In [24]:

```
alt.Chart(df3[df3.produto=='GASOLINA']).mark_rect().encode(
    y='bandeira:O',
    x='ano:O',
    color='venda:Q'
)
```

Out[24]:



Melhores Preços de Venda por Bandeira

In [32]:

```
def gerar_bumpchart_precos(produto='ETANOL'):

    base = df3[df3.produto==produto]
    #base = df[(df.produto=='ETANOL') & (df.bandeira.isin(['IPIRANGA', 'PETROBRAS'])
    source = base.groupby([pd.Grouper(key="data", freq="6M"), "bandeira"]).mean().r

    grafico = alt.Chart(source).mark_line(point=True).encode(
        x=alt.X("data:0", timeUnit="yearmonth", title="Data"),
        y="rank:0",
        color=alt.Color("bandeira:N")
    ).transform_window(
        rank="rank()",
        sort=[alt.SortField("venda", order="ascending")],
        groupby=["data"]
    ).properties(
        title="Melhores Preços de Venda por Bandeira - %s" % produto,
        width=600,
        height=300
    )

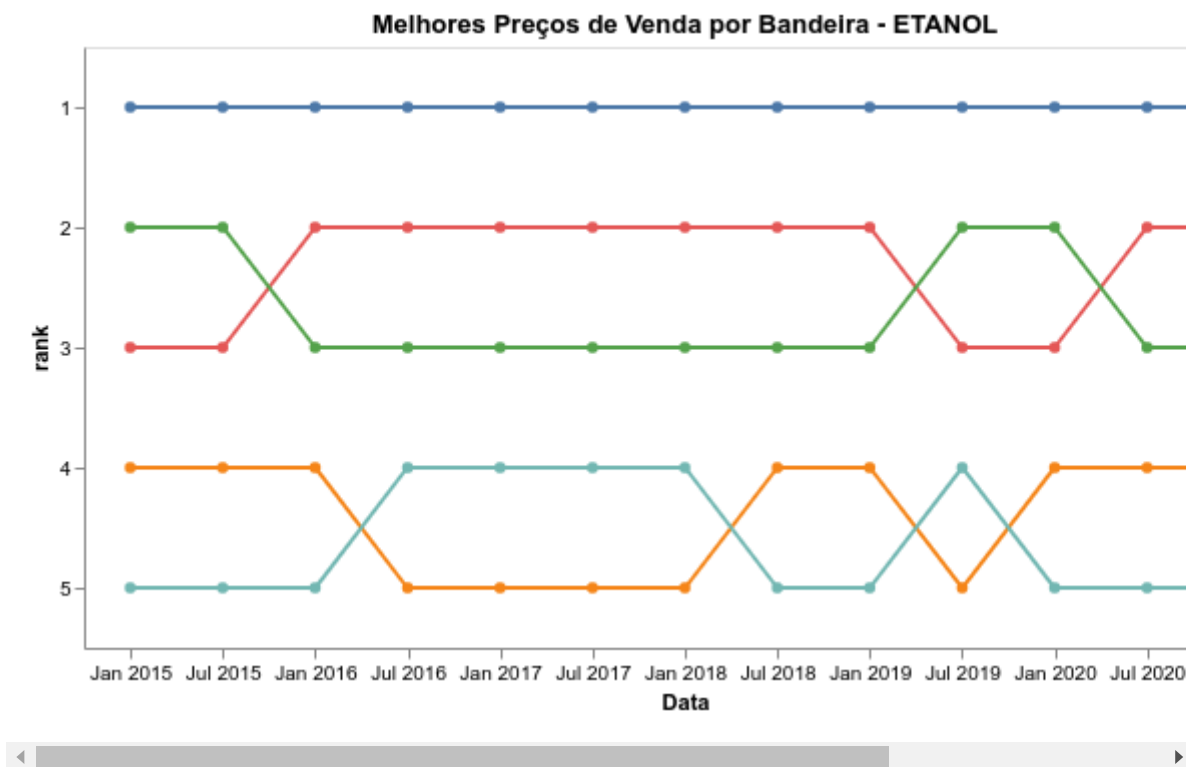
    return grafico
```

Melhores Preços de Venda de Etanol por Bandeira

In [33]:

```
gerar_bumpchart_precos(produto='ETANOL')
```

Out[33]:

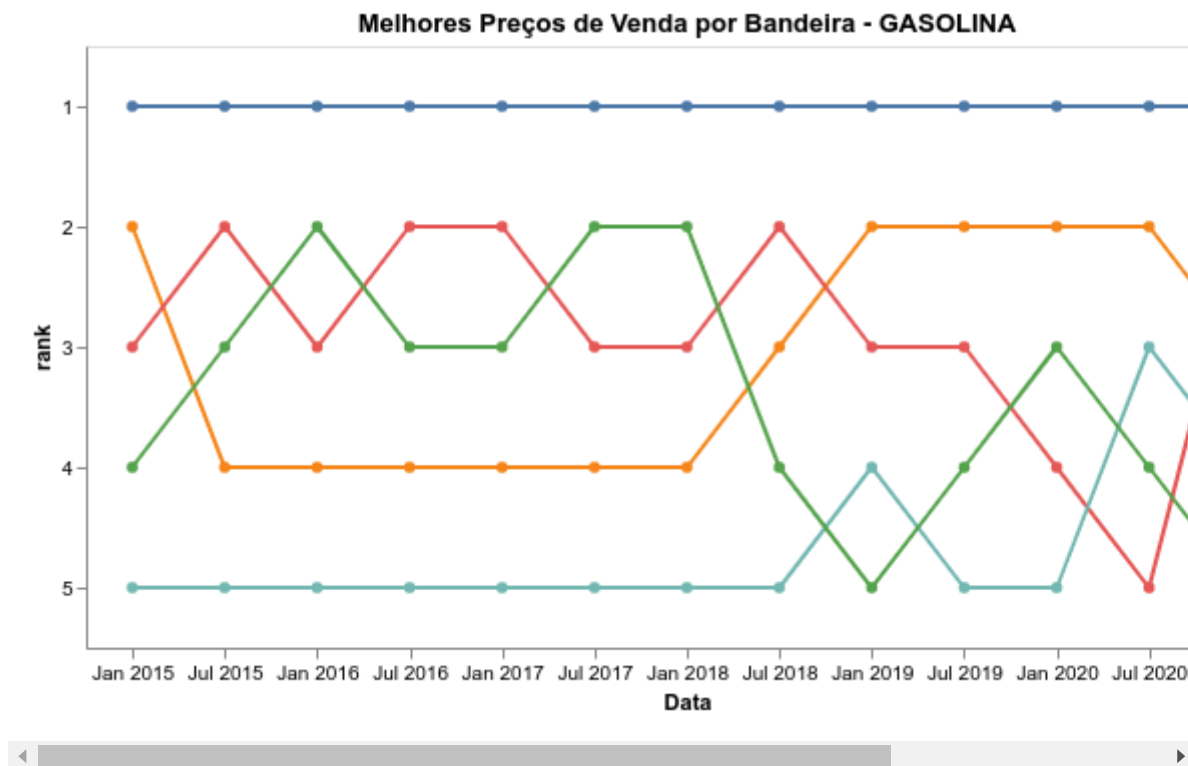


Melhores Preços de Venda de Gasolina por Bandeira

In [34]:

```
gerar_bumpchart_precos(produto='GASOLINA')
```

Out[34]:



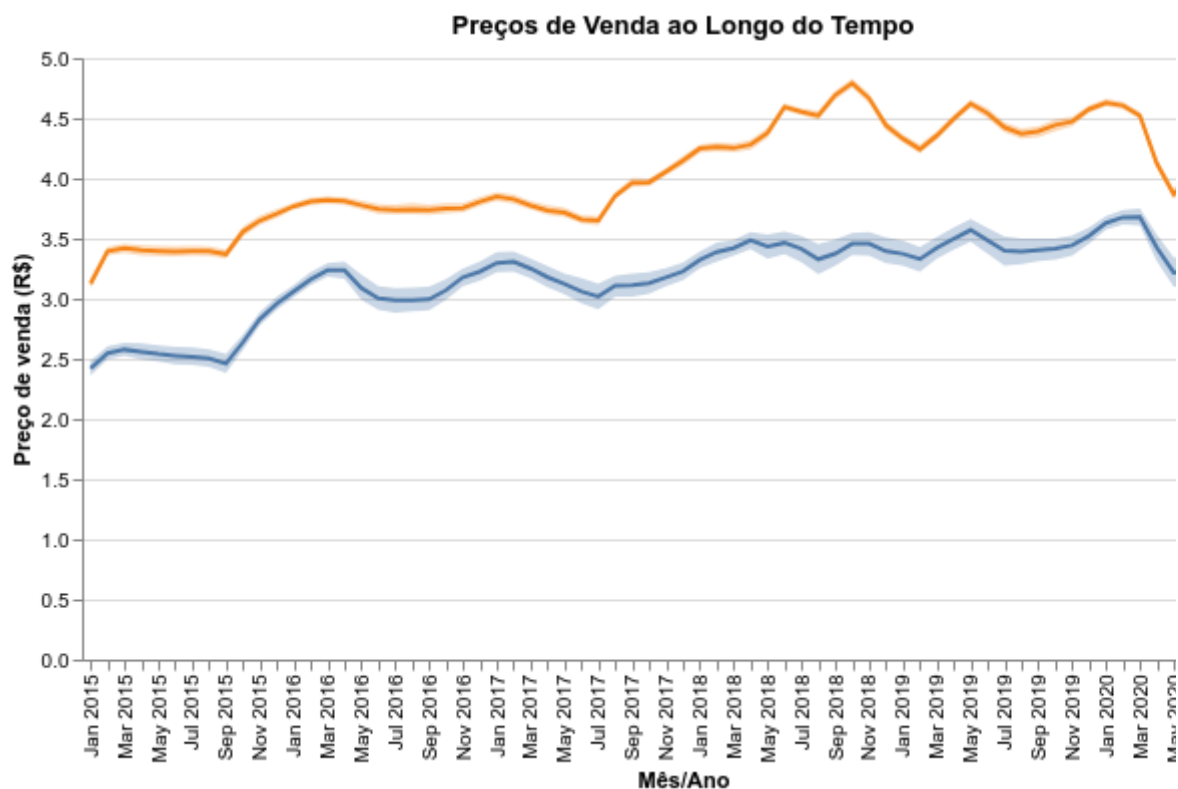
Preços de Venda ao Longo do Tempo

In [35]:

```
base = df3#[df.produto=='ETANOL']

line = alt.Chart(base).mark_line().encode(
    x=alt.X("data:0", timeUnit="yearmonth", title="Mês/Ano"),
    y='mean(venda)',
    color='produto'
)
band = alt.Chart(base).mark_errorband(extent='ci').encode(
    x=alt.X("data:0", timeUnit="yearmonth", title="Mês/Ano"),
    y=alt.Y('venda', title='Preço de venda (R$)'),
    color='produto'
)
(band + line).properties(
    title="Preços de Venda ao Longo do Tempo",
    width=600,
    height=300
)
```

Out[35]:



In []:

