

Pergunta 3. Em qual UF e período vale a pena colocar etanol invés de gasolina (considerando a eficiência do etanol a 75% da gasolina)?

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

Carregar dados

In [49]:

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

In [50]:

```
from funcoes_apoio import preparar_dataframe
```

In [130]:

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

Out[130]:

	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 [131]:

```
df2 = df.drop(['compra', 'margem', 'pmargem', 'bandeira', 'ams'], axis=1).\
    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[131]:

	regiao	uf	produto	ano	mes	venda
0	S	PR	GASOLINA	2015	1	3.014511
1	S	PR	GASOLINA	2015	2	3.289405
2	S	PR	GASOLINA	2015	3	3.328853
3	S	PR	GASOLINA	2015	4	3.295943
4	S	PR	GASOLINA	2015	5	3.297062
...
19434	SE	SP	ETANOL	2020	7	2.587264
19435	SE	SP	ETANOL	2020	8	2.603317
19437	SE	SP	ETANOL	2020	10	2.872609
19438	SE	SP	ETANOL	2020	11	2.972598
19439	SE	SP	ETANOL	2020	12	3.030175

3819 rows × 6 columns

Separar e mesclar dados

In [132]:

```
dfe = df2[df2.produto=='ETANOL'].drop('produto', axis=1).\
    set_index(['regiao', 'uf', 'ano', 'mes'])
dfe
```

Out[132]:

				venda
regiao	uf	ano	mes	
S	PR	2015	1	2.045657
			2	2.237970
			3	2.271349
			4	2.223547
			5	2.213576
...
SE	SP	2020	7	2.587264
			8	2.603317
			10	2.872609
			11	2.972598
			12	3.030175

1902 rows × 1 columns

In [133]:

```
dfg = df2[df2.produto=='GASOLINA'].drop('produto', axis=1).\
    set_index(['regiao', 'uf', 'ano', 'mes'])
dfg
```

Out[133]:

				venda
regiao	uf	ano	mes	
S	PR	2015	1	3.014511
			2	3.289405
			3	3.328853
			4	3.295943
			5	3.297062
...
SE	SP	2020	7	3.965204
			8	4.041624
			10	4.229716
			11	4.224165
			12	4.247449

1917 rows × 1 columns

In [134]:

```
df3 = dfg.merge(dfe, left_index=True, right_index=True, suffixes=('_gasolina', '_etanol'))
df3 = df3.reset_index()
df3
```

Out[134]:

	regiao	uf	ano	mes	venda_gasolina	venda_etanol
0	S	PR	2015	1	3.014511	2.045657
1	S	PR	2015	2	3.289405	2.237970
2	S	PR	2015	3	3.328853	2.271349
3	S	PR	2015	4	3.295943	2.223547
4	S	PR	2015	5	3.297062	2.213576
...
1897	SE	SP	2020	7	3.965204	2.587264
1898	SE	SP	2020	8	4.041624	2.603317
1899	SE	SP	2020	10	4.229716	2.872609
1900	SE	SP	2020	11	4.224165	2.972598
1901	SE	SP	2020	12	4.247449	3.030175

1902 rows × 6 columns

In [135]:

```
df3['ams'] = df3['ano'].astype('int') * 100 + df3['mes'].astype('int')
```

"para saber se o etanol mais é vantajoso ou não, basta multiplicar o valor do litro da gasolina por 0,7. Se o valor resultante for menor que o do litro do etanol, é melhor abastecer com gasolina. Se for maior, o etanol é a melhor opção."

<https://g1.globo.com/economia/noticia/2022/03/30/etanol-ou-gasolina-veja-em-quais-estados-abastecer-com-alcool-e-mais-vantajoso-e-como-calcular.ghtml> (<https://g1.globo.com/economia/noticia/2022/03/30/etanol-ou-gasolina-veja-em-quais-estados-abastecer-com-alcool-e-mais-vantajoso-e-como-calcular.ghtml>)

In [136]:

```
def melhor_opcao(valor_gasolina, valor_etanol):
    return 'Gasolina' if valor_gasolina * 0.75 < valor_etanol else 'Etanol'

print(melhor_opcao(1, 1))
print(melhor_opcao(1.25, 1))
print(melhor_opcao(1.5, 1))
print(melhor_opcao(2, 1))
```

Gasolina

Gasolina

Etanol

Etanol

In [137]:

```
df3['melhor'] = df3.apply(lambda x: melhor_opcao(x.venda_gasolina, x.venda_etanol),
df3['melhor'] = df3['melhor'].astype('category')
df3
```

Out[137]:

	regiao	uf	ano	mes	venda_gasolina	venda_etanol	ams	melhor
0	S	PR	2015	1	3.014511	2.045657	201501	Etanol
1	S	PR	2015	2	3.289405	2.237970	201502	Etanol
2	S	PR	2015	3	3.328853	2.271349	201503	Etanol
3	S	PR	2015	4	3.295943	2.223547	201504	Etanol
4	S	PR	2015	5	3.297062	2.213576	201505	Etanol
...
1897	SE	SP	2020	7	3.965204	2.587264	202007	Etanol
1898	SE	SP	2020	8	4.041624	2.603317	202008	Etanol
1899	SE	SP	2020	10	4.229716	2.872609	202010	Etanol
1900	SE	SP	2020	11	4.224165	2.972598	202011	Etanol
1901	SE	SP	2020	12	4.247449	3.030175	202012	Etanol

1902 rows × 8 columns

In [138]:

```
df3['relacao'] = df3['venda_etanol'] / df3['venda_gasolina']
df3
```

Out[138]:

	regiao	uf	ano	mes	venda_gasolina	venda_etanol	ams	melhor	relacao
0	S	PR	2015	1	3.014511	2.045657	201501	Etanol	0.678603
1	S	PR	2015	2	3.289405	2.237970	201502	Etanol	0.680357
2	S	PR	2015	3	3.328853	2.271349	201503	Etanol	0.682322
3	S	PR	2015	4	3.295943	2.223547	201504	Etanol	0.674632
4	S	PR	2015	5	3.297062	2.213576	201505	Etanol	0.671378
...
1897	SE	SP	2020	7	3.965204	2.587264	202007	Etanol	0.652492
1898	SE	SP	2020	8	4.041624	2.603317	202008	Etanol	0.644126
1899	SE	SP	2020	10	4.229716	2.872609	202010	Etanol	0.679149
1900	SE	SP	2020	11	4.224165	2.972598	202011	Etanol	0.703713
1901	SE	SP	2020	12	4.247449	3.030175	202012	Etanol	0.713410

1902 rows × 9 columns

In [160]:

```
df3.describe()
```

Out[160]:

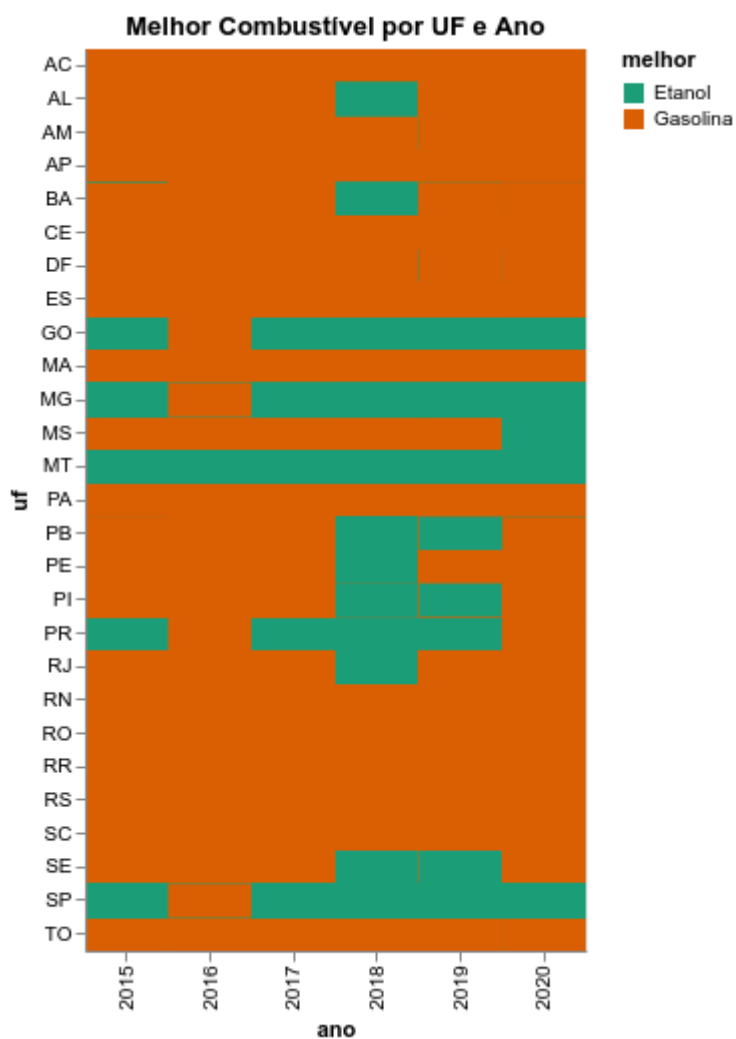
	ano	mes	venda_gasolina	venda_etanol	ams	relacao
count	1902.000000	1902.000000	1902.000000	1902.000000	1902.000000	1902.000000
mean	2017.460568	6.462145	4.050456	3.236127	201752.518927	0.800363
std	1.695764	3.464160	0.464130	0.458624	169.550991	0.085123
min	2015.000000	1.000000	2.910695	1.914087	201501.000000	0.556329
25%	2016.000000	3.000000	3.689958	2.918169	201606.000000	0.752513
50%	2017.000000	6.000000	4.013653	3.285212	201712.000000	0.804288
75%	2019.000000	10.000000	4.426309	3.583281	201906.000000	0.851465
max	2020.000000	12.000000	5.381083	4.448928	202012.000000	1.273496

Melhor Combustível por UF e Ano

In [143]:

```
alt.Chart(df3).mark_rect().encode(
    y='uf:0',
    x='ano:0',
    color=alt.Color('melhor', scale=alt.Scale(scheme='dark2'))
).properties(
    title="Melhor Combustível por UF e Ano",
    width=250,
    height=450
)
```

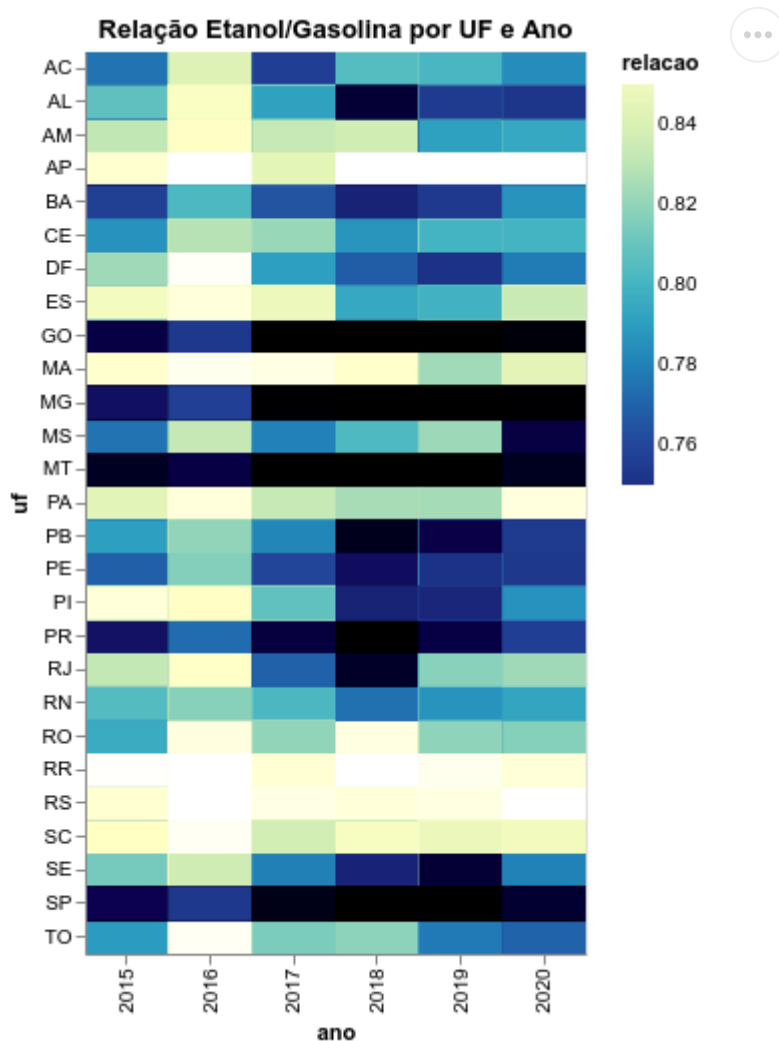
Out[143]:



In [176]:

```
alt.Chart(df3).mark_rect().encode(
    y='uf:0',
    x='ano:0',
    color=alt.Color('relacao', scale=alt.Scale(scheme='yellowgreenblue', domain=[0.
]).properties(
    title="Relação Etanol/Gasolina por UF e Ano",
    width=250,
    height=450
)
```

Out[176]:

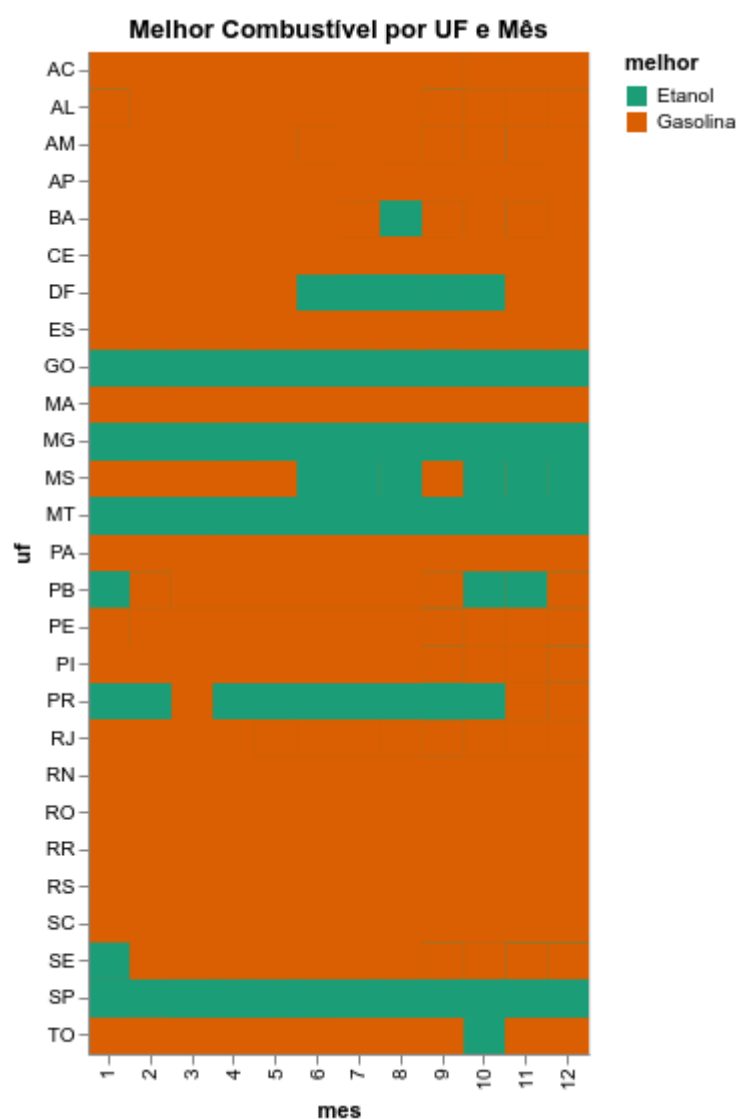


Melhor Combustível por UF e Mês

In [141]:

```
alt.Chart(df3).mark_rect().encode(
    y='uf:0',
    x='mes:0',
    color=alt.Color('melhor', scale=alt.Scale(scheme='dark2'))
).properties(
    title="Melhor Combustível por UF e Mês",
    width=250,
    height=500
)
```

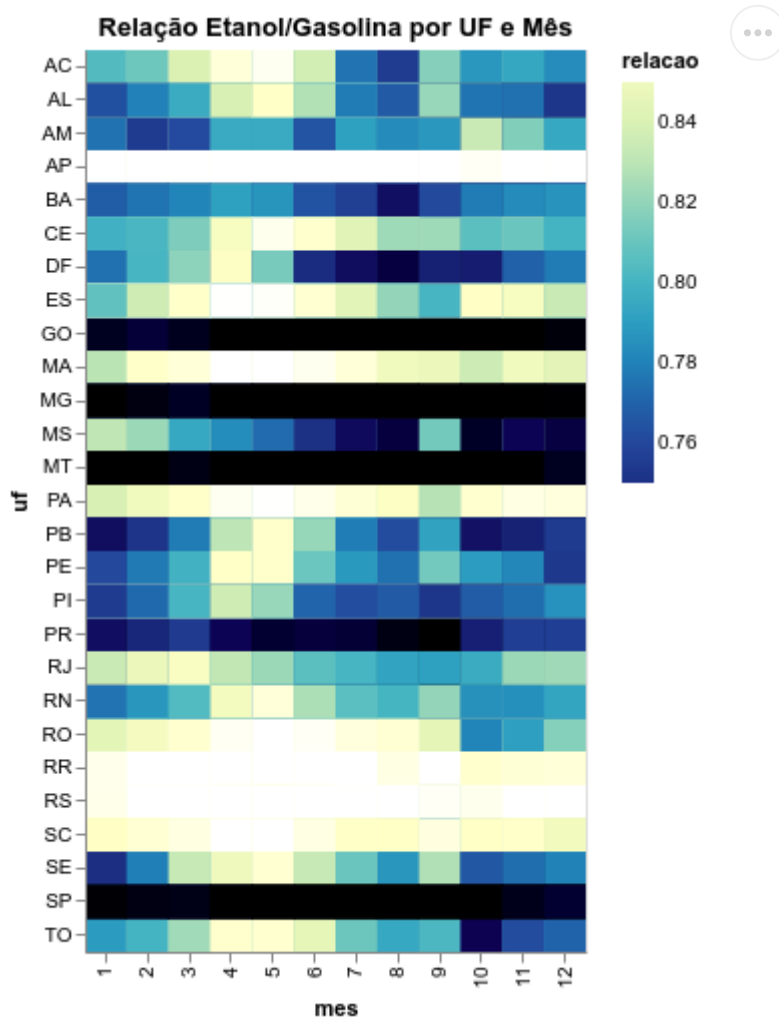
Out[141]:



In [177]:

```
alt.Chart(df3).mark_rect().encode(
    y='uf:O',
    x='mes:O',
    color=alt.Color('relacao', scale=alt.Scale(scheme='yellowgreenblue', domain=[0.
]).properties(
    title="Relação Etanol/Gasolina por UF e Mês",
    width=250,
    height=450
)
```

Out[177]:

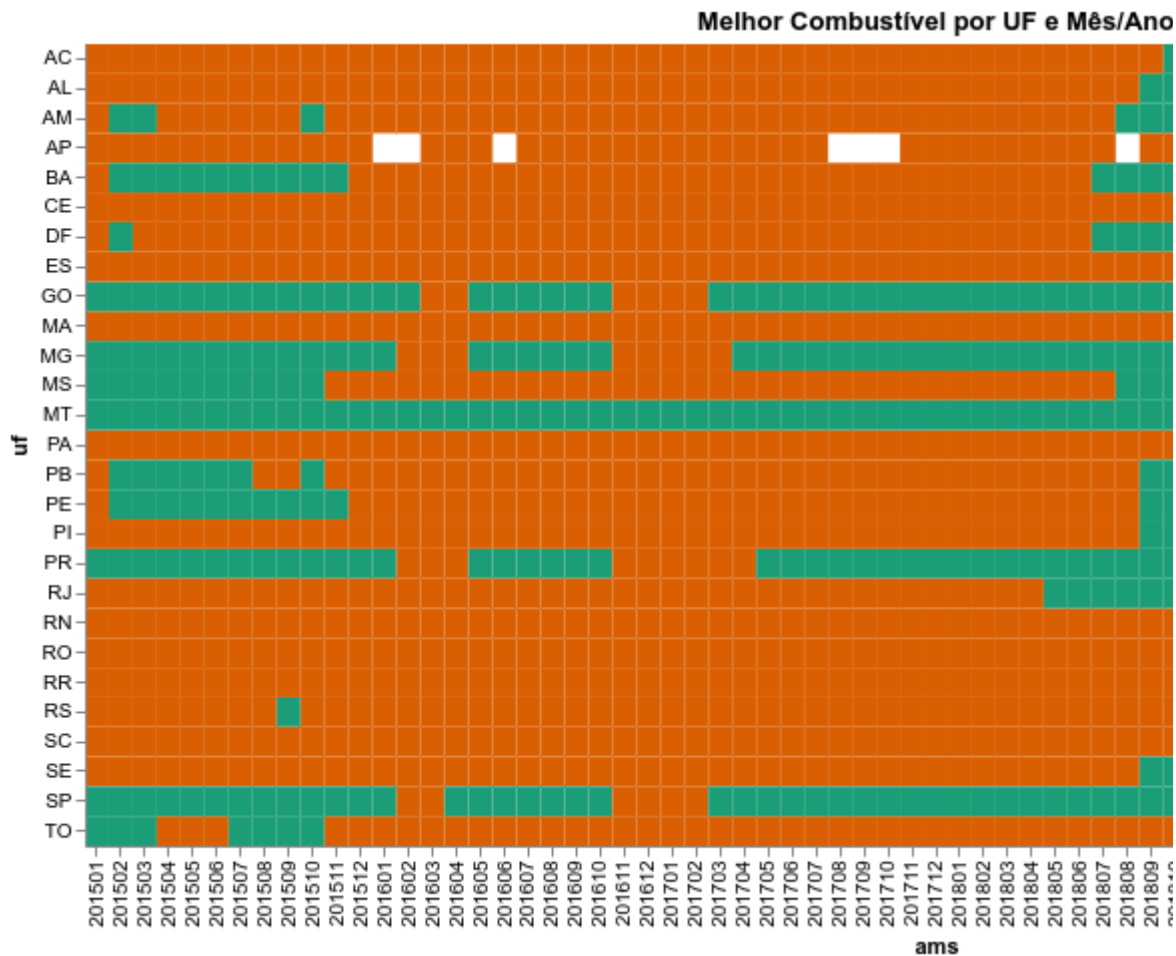


Melhor Combustível por UF e Mês/Ano

In [142]:

```
alt.Chart(df3).mark_rect().encode(
    y='uf:O',
    x='ams:O',
    color=alt.Color('melhor', scale=alt.Scale(scheme='dark2'))
).properties(
    title="Melhor Combustível por UF e Mês/Ano",
    width=850,
    height=400
)
```

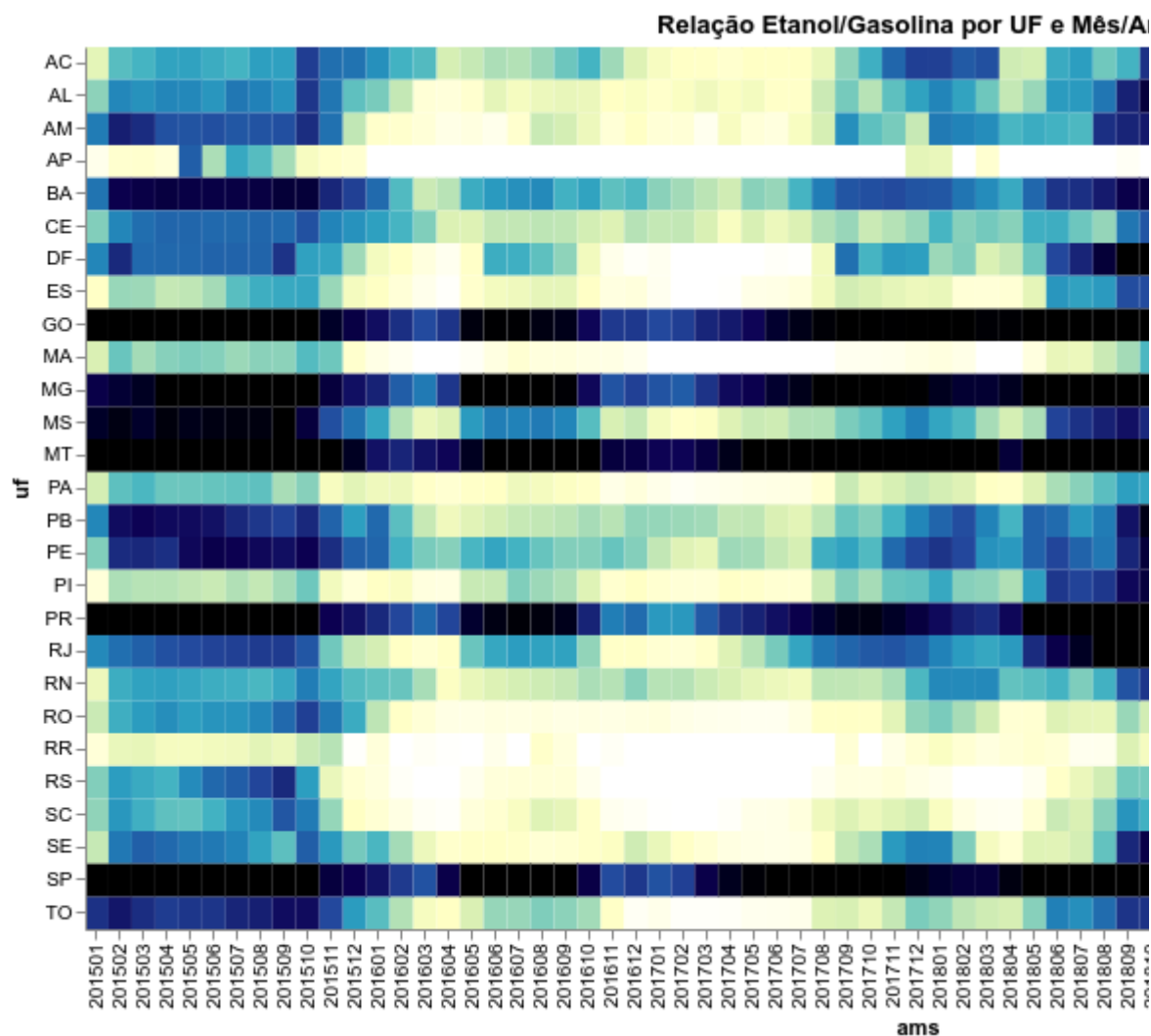
Out[142]:



In [180]:

```
alt.Chart(df3).mark_rect().encode(
    y='uf:0',
    x='ams:0',
    color=alt.Color('relacao', scale=alt.Scale(scheme='yellowgreenblue', domain=[0.
]).properties(
    title="Relação Etanol/Gasolina por UF e Mês/Ano",
    width=850,
    height=450
)
```

Out[180]:



In []: