

trabalho-isabella-enxuto

April 2, 2019

1 Trabalho de análise multivariada

Nome: Isabella Lucinda

Neste trabalho serão estudadas as produções agrícolas mais representativas nas 853 cidades do estado de Minas Gerais. Os dados foram retirados do site do SIDRA-IBGE do ano de 2017. Aqui foram selecionados o rendimento e o valor de 8 culturas: abacaxi, alho, batata inglesa, café arábica, feijão, laranja, milho e tomate, que são as culturas que se mostraram mais intensas no estado de Minas Gerais. Será analisado como essas culturas se correlacionam no quesito valor e rendimento, e os grupos de cidades com maior relação em cada uma dessas produções.

```
In [1]: import pandas as pd
import numpy as np
import scipy.stats as stats

%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
# sns.set(style='whitegrid', palette='pastel')

# para a análise de componentes principais
from sklearn.decomposition import PCA
from sklearn.preprocessing import scale
```

1.1 Análise de variâncias e correlações

```
In [2]: dados = pd.read_csv('/home/patricia/drive/unifal/analise-multivariada/2018-2/trabalho/al
```

```
In [3]: # dados2: incluir código para usar depois
dados2 = dados
# resultado2 = resultado.reset_index()
cod = pd.read_csv('https://patriciasiqueira.github.io/arquivos/codigos-mun.csv')
cod = cod[['mun', 'nome_mun', 'nome_meso']]
cod.rename(columns={'mun': 'ibge7'}, inplace=True)
dados2 = dados2.merge(cod)
```

```
In [4]: dados = dados.loc[:, ['ibge7', 'abacaxi_rend', 'alho_rend',
'batata_inglesa_rend', 'cafe_arab_rend', 'feijao_rend', 'laranja_rend',
'milho_rend', 'tomate_rend', 'abacaxi_val', 'alho_val',
```

```
'batata_inglesa_val', 'cafe_arab_val', 'feijao_val', 'laranja_val',
'milho_val', 'tomate_val']]
```

```
In [5]: dados.iloc[:,1:].describe()
```

```
Out[5]:
```

	abacaxi_rend	alho_rend	batata_inglesa_rend	cafe_arab_rend	\
count	853.000000	853.000000	853.000000	853.000000	
mean	2309.691676	662.023447	4284.329426	829.987104	
std	6743.568277	2492.143006	10644.628485	889.807051	
min	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	699.000000	
75%	0.000000	0.000000	0.000000	1482.000000	
max	40000.000000	18700.000000	47300.000000	4767.000000	

	feijao_rend	laranja_rend	milho_rend	tomate_rend	abacaxi_val	\
count	853.000000	853.000000	853.000000	853.000000	853.000000	
mean	930.028136	4725.558030	4006.121923	18622.960141	236.065651	
std	668.774271	8373.525184	2468.773486	29194.697804	2799.782545	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	438.000000	0.000000	1905.000000	0.000000	0.000000	
50%	760.000000	0.000000	3760.000000	0.000000	0.000000	
75%	1250.000000	7000.000000	6000.000000	47667.000000	0.000000	
max	3455.000000	49146.000000	11846.000000	130000.000000	56400.000000	

	alho_val	batata_inglesa_val	cafe_arab_val	feijao_val	\
count	853.000000	853.000000	853.000000	853.000000	
mean	393.813599	970.293083	12455.618992	1325.091442	
std	5817.052950	8554.222552	33364.517854	7550.418859	
min	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	27.000000	
50%	0.000000	0.000000	33.000000	128.000000	
75%	0.000000	0.000000	5751.000000	461.000000	
max	144233.000000	180702.000000	459021.000000	147075.000000	

	laranja_val	milho_val	tomate_val
count	853.000000	853.000000	853.000000
mean	554.731536	3735.719812	649.609613
std	4802.082110	12817.178522	3868.552148
min	0.000000	0.000000	0.000000
25%	0.000000	72.000000	0.000000
50%	0.000000	311.000000	0.000000
75%	12.000000	1572.000000	98.000000
max	74888.000000	161220.000000	73440.000000

As culturas que possuem a maior media de produção é a do valor do cafe e o rendimento do tomate. E as menores são o rendimento do alho e o valor do abacaxi. As culturas que apresentaram maior desvio padrão foram o valor do cafe e o rendimento do tomate. O cafe possuir maior media

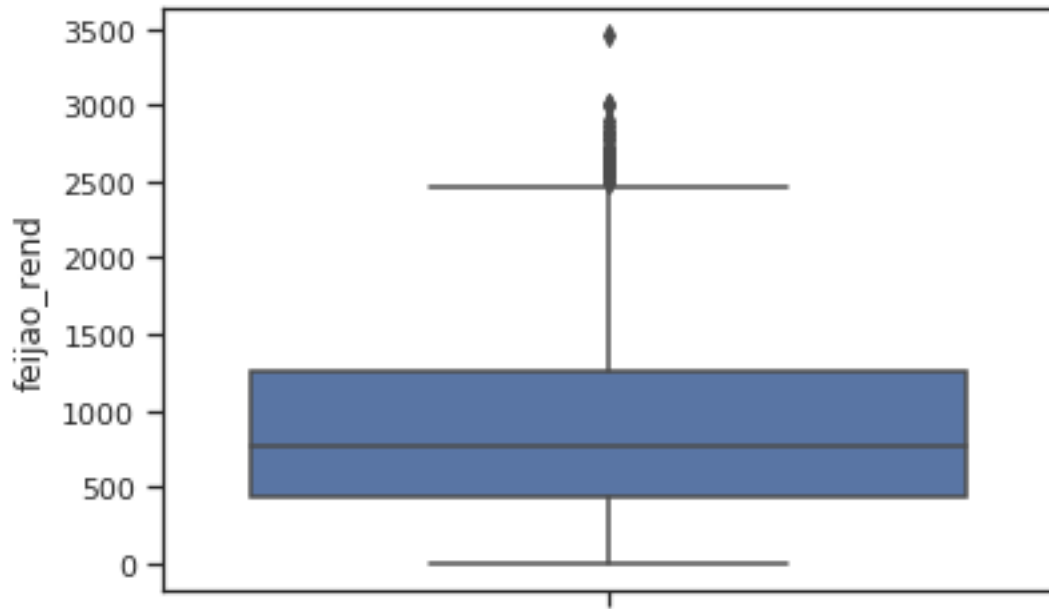
e desvio padrão pode ser explicado pelos quartis, que mostram que pelo menos metade do estado atua na produção de café. Porém, a cultura mais popular no estado é a de feijão, que por outro lado, não possui maior media nem desvio padrão. Até 75% dos municípios de Minas Gerais produzem somente café, feijão, laranja, milho e tomate; 50% produzem somente cafe, feijão e milho; e 25% produzem somente feijão e milho. Ou seja, a maior parte das cidades não produz grande parte das culturas analisadas.

```
In [29]: # médias
         dados.iloc[:,1:].mean()

Out[29]: abacaxi_rend      2309.691676
         alho_rend         662.023447
         batata_inglesa_rend 4284.329426
         cafe_arab_rend     829.987104
         feijao_rend       930.028136
         laranja_rend     4725.558030
         milho_rend        4006.121923
         tomate_rend      18622.960141
         abacaxi_val       236.065651
         alho_val          393.813599
         batata_inglesa_val  970.293083
         cafe_arab_val     12455.618992
         feijao_val        1325.091442
         laranja_val       554.731536
         milho_val         3735.719812
         tomate_val        649.609613
         grupo             0.570926
         dtype: float64
```

Se tratando de rendimento, a produção do tomate é a mais representativa em Minas. Já a de alho, é a menos representativa. Isso significa dizer que a produção de tomate por area plantada é a maior de Minas, e que a de alho não é muito sugestiva. Já se referindo ao valor da produção, o café arabica é de longe o mais representativo para o estado. E a produção de abacaxi é a de menos valor , seguida pela produção de alho. Ou seja, a pesar de o tomate ter maior rendimento que o cafe, o valor do cafe supera o valor do tomate.

```
In [34]: sns.boxplot(dados.feijao_rend, orient='v');
```



```
In [7]: dados.iloc[:,1:].corr()
```

```
Out[7]:
```

	abacaxi_rend	alho_rend	batata_inglesa_rend	\
abacaxi_rend	1.000000	0.074313	-0.040794	
alho_rend	0.074313	1.000000	0.275778	
batata_inglesa_rend	-0.040794	0.275778	1.000000	
cafe_arab_rend	0.003644	0.058885	0.233645	
feijao_rend	0.107599	0.189904	0.325380	
laranja_rend	0.181755	0.070168	0.065020	
milho_rend	0.022276	0.098561	0.390019	
tomate_rend	-0.024418	0.122430	0.193854	
abacaxi_val	0.348452	-0.019534	-0.028824	
alho_val	-0.014558	0.410023	0.186603	
batata_inglesa_val	0.026514	0.242100	0.341929	
cafe_arab_val	0.018280	0.103929	0.234886	
feijao_val	0.032094	0.092679	0.162653	
laranja_val	0.275771	-0.015003	0.035137	
milho_val	0.086857	0.293879	0.333044	
tomate_val	-0.005903	0.040028	0.136110	

	cafe_arab_rend	feijao_rend	laranja_rend	milho_rend	\
abacaxi_rend	0.003644	0.107599	0.181755	0.022276	
alho_rend	0.058885	0.189904	0.070168	0.098561	
batata_inglesa_rend	0.233645	0.325380	0.065020	0.390019	
cafe_arab_rend	1.000000	0.225841	0.160290	0.278798	
feijao_rend	0.225841	1.000000	0.168932	0.588867	

laranja_rend	0.160290	0.168932	1.000000	0.158321
milho_rend	0.278798	0.588867	0.158321	1.000000
tomate_rend	0.184871	0.145681	0.175245	0.227976
abacaxi_val	0.003227	0.083616	0.117482	0.058536
alho_val	0.049858	0.109334	-0.028401	0.123182
batata_inglesa_val	0.076731	0.225866	0.044349	0.178646
cafe_arab_val	0.356915	0.198202	0.190280	0.271992
feijao_val	0.158329	0.299607	0.101018	0.176495
laranja_val	0.062210	0.071130	0.313967	0.096139
milho_val	0.248514	0.426703	0.195706	0.352321
tomate_val	0.166837	0.148216	0.122054	0.134926

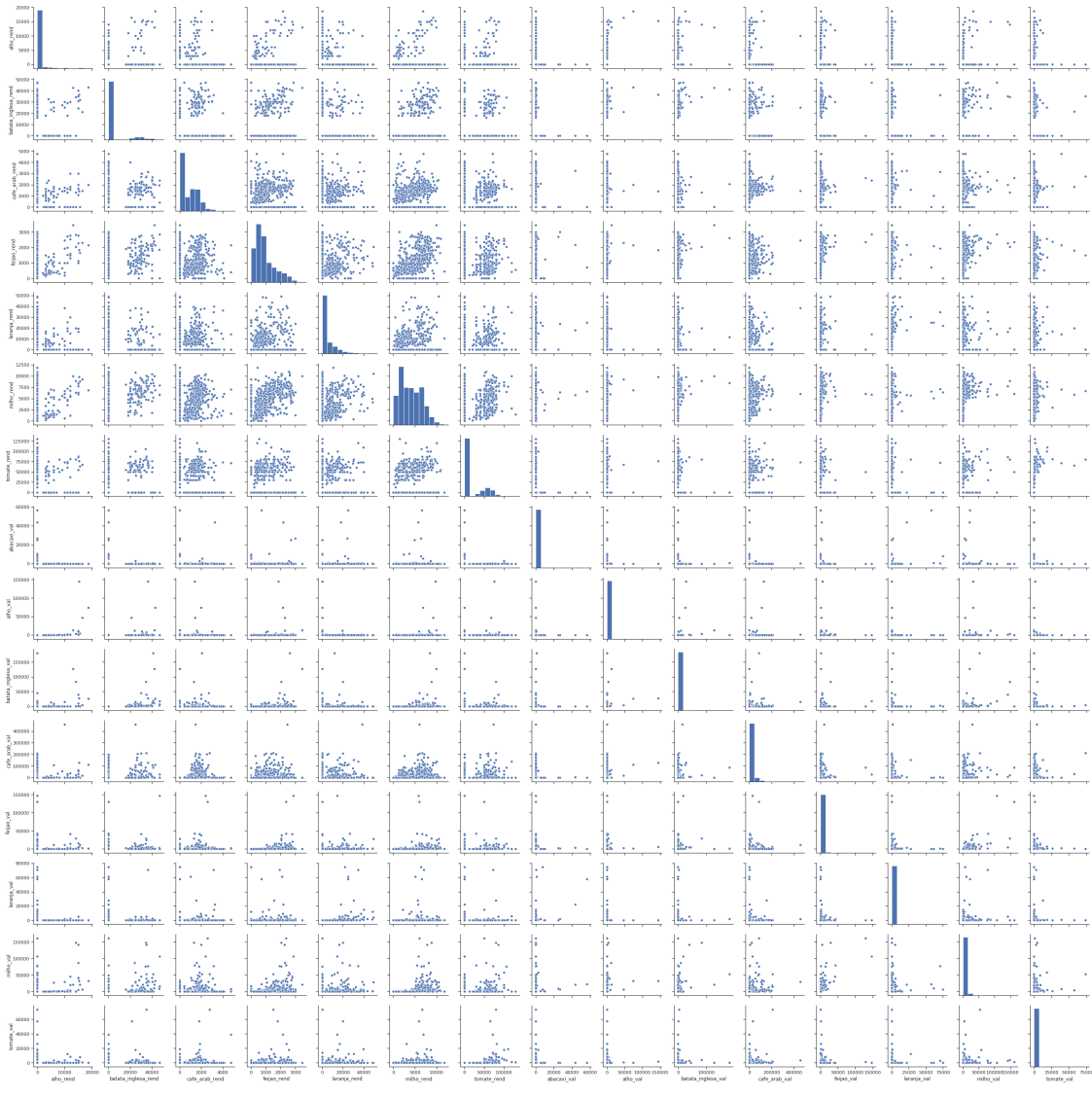
	tomate_rend	abacaxi_val	alho_val	batata_inglesa_val	\
abacaxi_rend	-0.024418	0.348452	-0.014558	0.026514	
alho_rend	0.122430	-0.019534	0.410023	0.242100	
batata_inglesa_rend	0.193854	-0.028824	0.186603	0.341929	
cafe_arab_rend	0.184871	0.003227	0.049858	0.076731	
feijao_rend	0.145681	0.083616	0.109334	0.225866	
laranja_rend	0.175245	0.117482	-0.028401	0.044349	
milho_rend	0.227976	0.058536	0.123182	0.178646	
tomate_rend	1.000000	-0.046816	0.069254	0.033252	
abacaxi_val	-0.046816	1.000000	-0.005502	-0.002456	
alho_val	0.069254	-0.005502	1.000000	0.183833	
batata_inglesa_val	0.033252	-0.002456	0.183833	1.000000	
cafe_arab_val	0.217636	-0.025794	0.152807	0.120070	
feijao_val	0.055616	0.001610	0.026445	0.101040	
laranja_val	0.018806	0.421886	-0.007097	0.008363	
milho_val	0.203167	0.071285	0.133844	0.373033	
tomate_val	0.319682	-0.008204	0.007678	0.019831	

	cafe_arab_val	feijao_val	laranja_val	milho_val	\
abacaxi_rend	0.018280	0.032094	0.275771	0.086857	
alho_rend	0.103929	0.092679	-0.015003	0.293879	
batata_inglesa_rend	0.234886	0.162653	0.035137	0.333044	
cafe_arab_rend	0.356915	0.158329	0.062210	0.248514	
feijao_rend	0.198202	0.299607	0.071130	0.426703	
laranja_rend	0.190280	0.101018	0.313967	0.195706	
milho_rend	0.271992	0.176495	0.096139	0.352321	
tomate_rend	0.217636	0.055616	0.018806	0.203167	
abacaxi_val	-0.025794	0.001610	0.421886	0.071285	
alho_val	0.152807	0.026445	-0.007097	0.133844	
batata_inglesa_val	0.120070	0.101040	0.008363	0.373033	
cafe_arab_val	1.000000	0.121708	0.018686	0.268237	
feijao_val	0.121708	1.000000	0.004463	0.679858	
laranja_val	0.018686	0.004463	1.000000	0.172494	
milho_val	0.268237	0.679858	0.172494	1.000000	
tomate_val	0.200468	0.025791	0.023214	0.176438	

	tomate_val
abacaxi_rend	-0.005903
alho_rend	0.040028
batata_inglesa_rend	0.136110
cafe_arab_rend	0.166837
feijao_rend	0.148216
laranja_rend	0.122054
milho_rend	0.134926
tomate_rend	0.319682
abacaxi_val	-0.008204
alho_val	0.007678
batata_inglesa_val	0.019831
cafe_arab_val	0.200468
feijao_val	0.025791
laranja_val	0.023214
milho_val	0.176438
tomate_val	1.000000

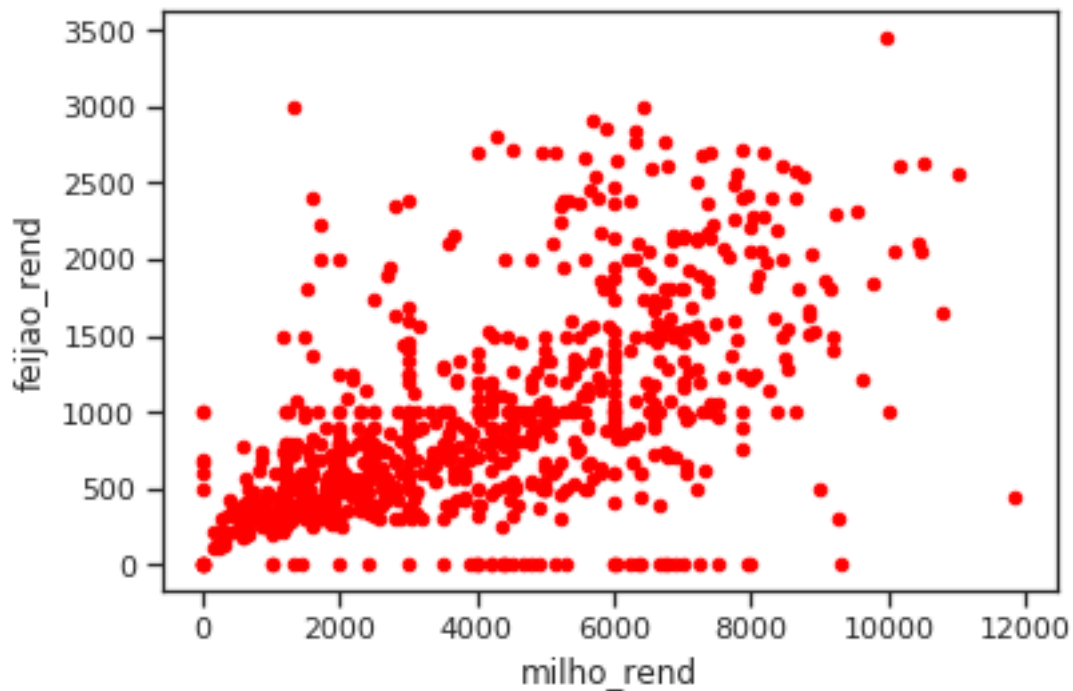
As culturas que possuem maior correlação positiva são o feijão e o milho, em torno de 0,67, o que significa dizer que, segundo os dados, nas regiões de Minas onde se concentra maior produção de feijão também se concentram as de milho. Apesar de muitas culturas possuírem correlação positiva baixa, as menores apresentadas foram abacaxi e café, em torno de 0,003. Isso significa que grande parte das agriculturas apresentadas não possuem relação quanto as cidades de plantio no estado. Existem também as culturas que apresentam correlação negativa, mas também são muito pequenas e não muito sugestivas, variam de -0,005 a -0,025.

```
In [8]: sns.set(style='ticks')
        sns.pairplot(dados.iloc[:,2:]);
```



Os graficos esboçam o que foi apresentado na tabela, mostrando a forte correlação entre milho e feijão, tanto no valor quanto no rendimento. E também a relação fraca entre as produções agrícolas dentro do estado de Minas.

```
In [9]: dados.plot.scatter('milho_rend', 'feijao_rend', s=20, c='red'); # há opções (usar SHL
```



1.2 ACP usando a matriz de correlações

In [10]: *# subselecionar variáveis*

```
X = dados.iloc[:,1:]
```

```
# efetua a ACP
```

```
pca = PCA()
```

```
resultado_pca = pca.fit_transform(scale(X)) # scale(X) padroniza os dados, como se us
```

```
resultado = pd.DataFrame({'cp1':resultado_pca[:, 0], 'cp2':resultado_pca[:, 1]}, index=
```

/home/patricia/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:6: DataConversionWarn

As variáveis apresentam 16 componentes principais.

Y_1 : As variáveis que mais influenciam o CP1 são nas produções de milho, feijão e café de forma positiva. Assim, quanto maior o valor do escore dos municípios do CP1 maior a produção dessas culturas em Minas. Todos os coeficientes do CP1 são positivos, ou seja, quanto maior o escore do município maior o valor de todas as variáveis, com maior predominância na produção de milho, feijão e café.

Y_2 : os maiores coeficientes, positivos, são os das produções de abacaxi e laranja. Ou seja, quanto maior o escore do CP2, mais alta será a produção de abacaxi e laranja nas cidades de Minas.

Y_3 : o terceiro componente é mais influenciado positivamente pela produção de alho, e negativamente pelo tomate. Ou seja, quando próximos de zero, eles se anulam, e quando o escore de um estiver mais alto, do outro estará mais baixo.

Seriam necessários pelo menos 8 componentes, que juntos explicam 73,3% da variância total. Porém, como 8 componentes seriam difíceis de analisar, ficaremos com 3, que juntos explicam 42,49% da variância total.

Podemos observar as cidades, como por exemplo, Sacramento, Paracatu, Patrocínio, Uberaba, que possuem forte relação como CP1, ou seja, são cidades fortes na produção de milho feijão e café. Também podemos observar as cidades como Frutal, Monte Alegre de Minas, que possuem altos escores no CP2, o que significa que são cidades fortes na produção de abacaxi e laranja. Observa-se cidades como Uberlândia que possui escores relativamente altos para os dois componentes.

```
In [11]: # variâncias acumuladas
tot = sum(pca.explained_variance_)
var_exp = [(i / tot)*100 for i in sorted(pca.explained_variance_, reverse=True)]
cum_var_exp = np.cumsum(var_exp)
# número de CPs gerados (k = p)
cp = ['CP'+str(i+1) for i in range(len(pca.components_))]
# dataframe com as variâncias acumuladas para criar o scree plot
df = pd.DataFrame({'cp': cp, 'var_exp': var_exp, 'cum_var_exp': cum_var_exp})
#df
```

1.3 Análise de Agrupamento

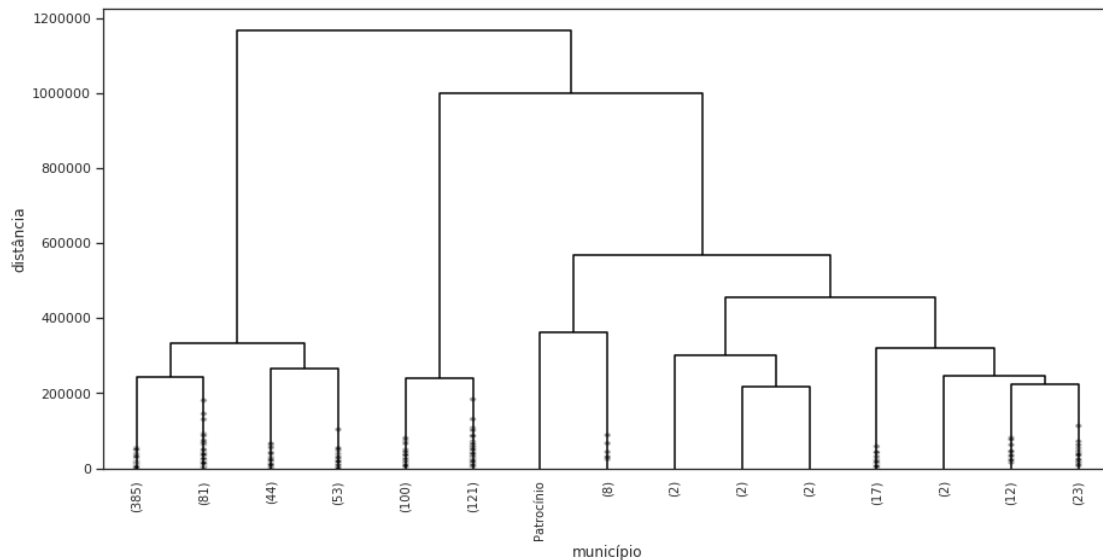
```
In [12]: from scipy.cluster.hierarchy import dendrogram, linkage
from scipy.spatial.distance import pdist
from scipy.cluster.hierarchy import cut_tree
from sklearn.cluster import KMeans
```

```
In [13]: Z = linkage(X, method='ward')
```

```
In [14]: max_d = 0
grupos = cut_tree(Z, height=max_d)
```

```
In [15]: fig, ax = plt.subplots(figsize=(15, 7))
ax = dendrogram(
    Z,
    truncate_mode='lastp', # mostrar apenas os p últimos grupos formados
    p=15, # quantos passos mostrar
    show_leaf_counts=True, # mostrar quantas observações há em cada grupo entre parênteses
    leaf_rotation=90., # rotação
    leaf_font_size=10., # tamanho da fonte
    labels=dados.index, # rótulos do eixo x
    show_contracted=True, # to get a distribution impression in truncated branches,
    above_threshold_color='black',
    color_threshold=0.1, # para que todas as linhas sejam da mesma cor
    # color_threshold=max_d, # para que os grupos fiquem com cores diferentes
)
plt.axhline(y=max_d, c='grey', lw=1, linestyle='dashed')
```

```
plt.xlabel('município')
plt.ylabel('distância');
```



Método escolhido:

```
In [16]: Z = linkage(X, method='ward')
# definir a distância de corte baseado no dendrograma
max_d = 800000
grupos = cut_tree(Z, height=max_d)
# incluir no resultado dos escores dos dois primeiros CPs a informação sobre os grupos
resultado['grupo'] = grupos
# contagem de observações em cada grupo
resultado.grupo.value_counts()
```

```
Out[16]: 1    563
         0    221
         2     69
         Name: grupo, dtype: int64
```

```
In [17]: # incluir no dataframe de dados as informações sobre a qual grupo cada observação pertence
dados['grupo'] = grupos
```

```
In [18]: # média dos grupos - todas as variáveis
# inclusive as não utilizadas para agrupar
dados.groupby('grupo').mean()
```

```
Out[18]:
```

	ibge7	abacaxi_rend	alho_rend	batata_inglesa_rend	\
grupo					
0	3.136194e+06	1964.705882	855.004525	5103.877828	
1	3.137078e+06	2457.312611	434.724689	2811.596803	

```

2      3.138718e+06    2210.144928    1898.550725          13676.043478

      cafe_arab_rend    feijao_rend    laranja_rend    milho_rend    tomate_rend  \
grupo
0      926.977376      946.674208      5978.990950    4261.552036    57994.683258
1      687.255773      871.735346      3775.509769    3638.781528      181.172291
2      1683.942029    1352.347826      8462.782609    6185.289855    42993.623188

      abacaxi_val      alho_val    batata_inglesa_val    cafe_arab_val  \
grupo
0      11.891403      251.846154          389.321267      5731.990950
1      346.003552      45.680284          283.264654      4488.811723
2      57.043478     3689.086957          8436.840580     98995.246377

      feijao_val    laranja_val      milho_val      tomate_val
grupo
0      859.642534    147.755656      2248.180995    1701.561086
1      849.497336    687.824156      2164.923623      5.065719
2     6696.449275    772.275362     21316.942029    2539.420290

```

```

In [19]: # mediana das variáveis para cada grupo
dados.groupby('grupo').median()

```

```

Out[19]:      ibge7    abacaxi_rend    alho_rend    batata_inglesa_rend    cafe_arab_rend  \
grupo
0      3135100          0          0          0          933
1      3137304          0          0          0          0
2      3143104          0          0          0         1680

      feijao_rend    laranja_rend    milho_rend    tomate_rend    abacaxi_val  \
grupo
0          800          0          4000          60000          0
1          680          0          3111          0          0
2         1289         5000          6519          51000          0

      alho_val    batata_inglesa_val    cafe_arab_val    feijao_val    laranja_val  \
grupo
0          0          0          168          189          0
1          0          0          0          90          0
2          0          0         81972          613          5

      milho_val    tomate_val
grupo
0          449          346
1          180          0
2         5544          249

```

O metodo de ward, por ser um metodo que apresenta resultados mais balanceados, é mais adequado para a analise feita neste trabalho.

grupo 0: o grupo 0 apresenta cidades com media e mediana altas na produção de tomate. Ou seja, fazem parte desse grupo os municipios com maiores valores de rendimento de tomate.

grupo 1: o grupo 1 apresenta media alta na produção de abacaxi, e fica em segundo na media dos municipios com valores mais altos de laranja. Isso significa dizer que, fazem parte do grupo 1 as cidades com maior representação na produção de frutas.

grupo 2: o grupo 2 é o que apresenta maior media e mediana para a maioria das culturas. Possui a media mais alta de municipios na produção de alho, batata inglesa, café, feijão, laranja e milho. É o grupo com o menor numero de cidades, porem as cidades que apresentam maiores participação na produção da maior parte das culturas de Minas.

1.3.1 As observações de cada grupo

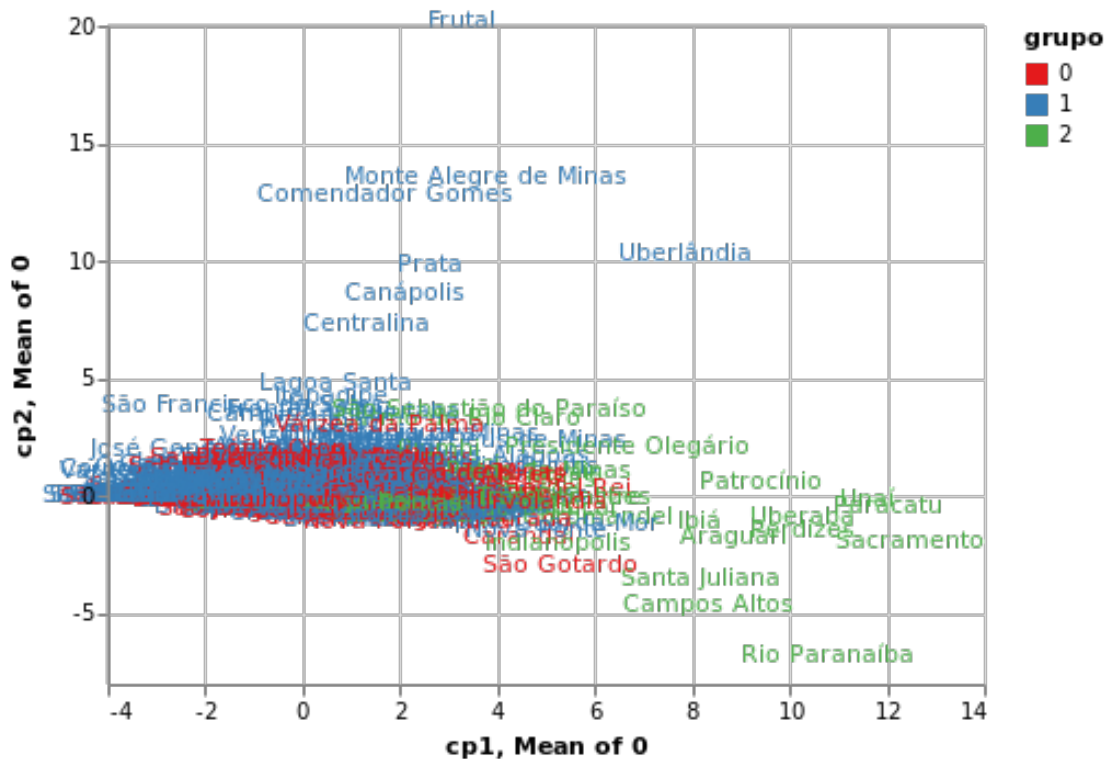
```
In [20]: resultado['município'] = resultado.index
```

```
In [21]: # pacote altair para fazer outros tipos de gráficos
# usaremos para o diagrama de dispersão com a cor
import altair as alt
```

```
In [22]: # gráficos
g = alt.Chart(resultado).mark_text().encode(
    alt.X('cp1', scale=alt.Scale(domain=[resultado.cp1.min(), resultado.cp1.max()])),
    alt.Y('cp2', scale=alt.Scale(domain=[resultado.cp2.min(), resultado.cp2.max()])),
    text='município',
    color=alt.Color('grupo:0', scale=alt.Scale(scheme='set1'))
)
# eixo x = 0
bar_x = alt.Chart(resultado).mark_rule(color='gray').encode(
    x='mean(0):Q'
)
# eixo y = 0
bar_y = alt.Chart(resultado).mark_rule(color='gray').encode(
    y='mean(0):Q'
)

g + bar_x + bar_y
```

```
Out[22]:
```



Com esse grafico se torna ainda mais facil a vizualização dos grupos. Patrocinio, como se destacou no calculo por outros metodos, pode-se observar que faz parte do grupo 2, que possui a maior media de municipios na maior parte das produções estudadas, e tambem possui uma alta relação com CP1.

1.4 K-Médias

In [23]: $k = 3$

```
kmeans = KMeans(n_clusters=k, random_state=10).fit(X)
# incluir no resultado dos escores dos dois primeiros CPs a informação sobre os grupos
# com o método das k-médias
resultado['grupo'] = kmeans.labels_
# contagens
resultado.grupo.value_counts()
```

```
Out[23]: 0      589
         2      223
         1       41
         Name: grupo, dtype: int64
```

```
In [24]: # incluir no dataframe de dados as informações sobre a qual grupo cada município pertence
dados['grupo'] = kmeans.labels_
```

```
In [25]: # média dos grupos - todas as variáveis
# inclusive as não utilizadas para agrupar
dados.groupby('grupo').mean()
```

```
Out[25]:
```

	ibge7	abacaxi_rend	alho_rend	batata_inglesa_rend	\
grupo					
0	3.137155e+06	2430.334465	469.864177	2808.847199	
1	3.136051e+06	3109.756098	1585.365854	13343.390244	
2	3.136695e+06	1843.946188	999.802691	6515.887892	

	cafe_arab_rend	feijao_rend	laranja_rend	milho_rend	tomate_rend	\
grupo						
0	692.084890	868.351443	3771.290323	3606.139219	935.483871	
1	1738.390244	1410.756098	10415.073171	6507.707317	40720.975610	
2	1027.206278	1004.547085	6199.968610	4602.645740	61277.242152	

	abacaxi_val	alho_val	batata_inglesa_val	cafe_arab_val	\
grupo					
0	330.404075	67.271647	519.760611	5048.573854	
1	28.926829	5608.439024	7252.536585	129820.707317	
2	24.977578	297.551570	1005.233184	10441.183857	

	feijao_val	laranja_val	milho_val	tomate_val
grupo				
0	1005.568761	536.059423	2083.174873	2.417657
1	6179.829268	1110.951220	20295.243902	3043.829268
2	1276.457399	501.784753	5055.937220	1918.816143

```
In [26]: # mediana das variáveis para cada grupo
dados.groupby('grupo').median()
```

```
Out[26]:
```

	ibge7	abacaxi_rend	alho_rend	batata_inglesa_rend	cafe_arab_rend	\
grupo						
0	3137601	0	0	0	333	
1	3139003	0	0	0	1735	
2	3136702	0	0	0	1164	

	feijao_rend	laranja_rend	milho_rend	tomate_rend	abacaxi_val	\
grupo						
0	680	0	3033	0	0	
1	1374	6000	6805	50000	0	
2	900	0	4400	60000	0	

	alho_val	batata_inglesa_val	cafe_arab_val	feijao_val	laranja_val	\
grupo						
0	0	0	7	93	0	
1	0	0	111376	936	5	
2	0	0	712	205	0	

	milho_val	tomate_val
grupo		
0	182	0
1	6709	225
2	597	400

Pelo metodo das K-Medias o grupo 0 é o grupo que apresenta as menores medias, ou seja, é o grupo que possui maior parte das cidades com baixa produção das culturas analisadas. Só se destaca no valor do abacaxi, possuindo a maior media, o que indica que os municipios que fazem parte deste grupo, possuem uma media de valores relativamente altos na cultura de abacaxi. O grupo 1 se destaca por possuir as maiores medias apresentadas. O que significa que fazem parte deste grupo as cidades que são mais ativas na produção de alho, batata inglesa, feijão, cafe, laranja e milho. Possuindo os maiores rendimento e valor de quase todas as culturas. O grupo 2 pode ser considerado como o grupo intermediario, nele se encontram as cidades que se destacam na produção de tomate, e que tambem sao fortes na produção das outras culturas, não tanto quanto o grupo 1, porem superior ao grupo 0.

Comparando a analise pelo metodo de ward e pela k-medias, se nota que os grupos, apesar de estarem invertidos, apresentam basicamente a mesma separação.

```
In [27]: resultado['município'] = resultado.index
```

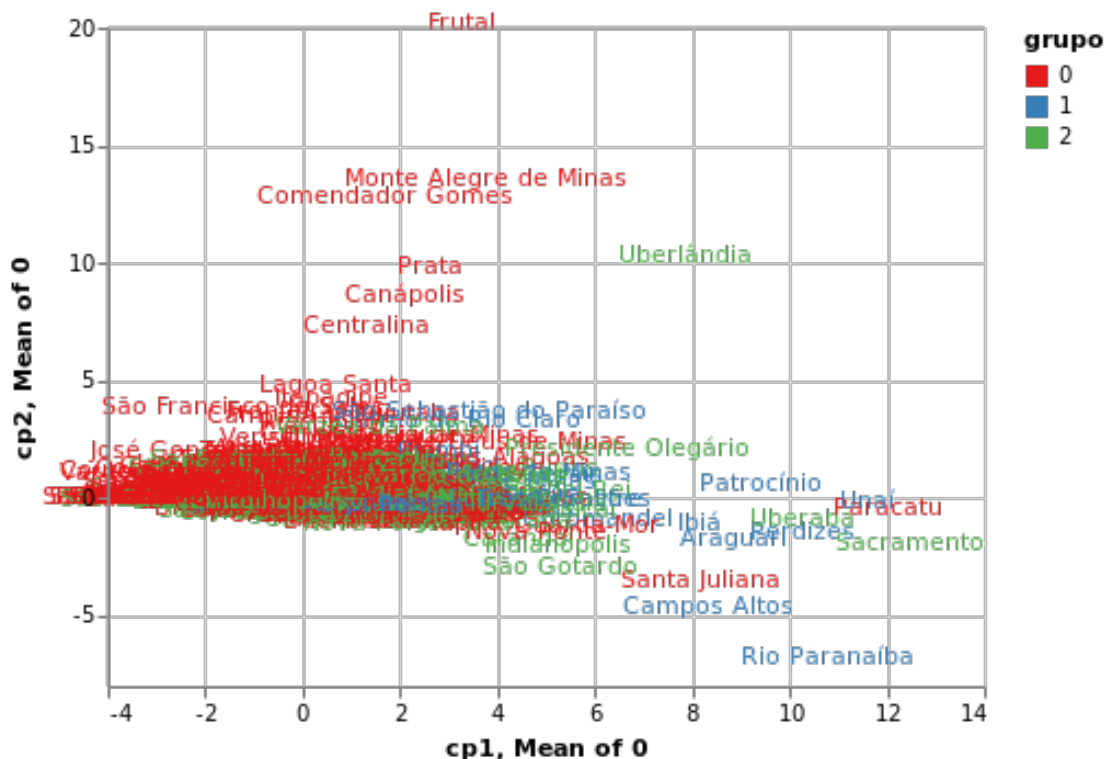
```
In [28]: # gráficos
```

```
g = alt.Chart(resultado).mark_text().encode(
    alt.X('cp1', scale=alt.Scale(domain=[resultado.cp1.min(), resultado.cp1.max()])),
    alt.Y('cp2', scale=alt.Scale(domain=[resultado.cp2.min(), resultado.cp2.max()])),
    text='município',
    color=alt.Color('grupo:0', scale=alt.Scale(scheme='set1'))
)
# eixo x = 0
bar_x = alt.Chart(resultado).mark_rule(color='gray').encode(
    x='mean(0):Q'
)

# eixo y = 0
bar_y = alt.Chart(resultado).mark_rule(color='gray').encode(
    y='mean(0):Q'
)

g + bar_x + bar_y
```

```
Out[28]:
```



Pelo metodo das K-medias, se observa que maior parte do grupo 0 possui relação mais forte com o CP2, e a maior parte do grupo 1, uma relação muito forte com CP1 e baixa com o CP2. Da mesma forma que observada no metodo de ward, o grupo que possui as cidades com maiores medias das produções de cafe, feijão, milho, etc, é o grupo que apresenta maior relação com o CP1, que possui escores fortes na produção de cafe, feijão e milho. E as cidades que possuem medias mais altas na produção de abacaxi, estão muito relacionadas com o CP2, que apresenta altos escores na produção de abacaxi e laranja.

1.5 Conclusão

Conclui-se então que, a pesar de serem as culturas que possuem maior representatividade em Minas Gerais, são poucos municípios que atuam, de fato, em suas produções. A análise de componentes principais, junto com a análise de agrupamento, mostraram as separações dos municípios de Minas de acordo com sua produção mais ativa. Por exemplo, cidades como Patrocínio, que apresenta uma das maiores relações com o CP1, é um município que se mostrou muito ativo na produção de culturas como, feijão, café, milho, etc. E também, cidades como Frutal, que é a cidade com maior relação com o CP2, que se mostrou um município muito ativo na produção de abacaxi.