main_py

September 15, 2021

1 Pie chart of samples

```python
[1]: import pandas as pd
     import matplotlib.pyplot as plt
```

```python
[2]: %matplotlib inline
```

```python
[3]: df = pd.read_csv("../../input/phenotypes/merged/_m/merged_phenotypes.csv",␣
     ↪index_col=0)
     df = df[df["Race"].isin(["AA", "CAUC"]) & (df["Dx"].isin(["Control",␣
     ↪"Schizo"]))].copy()
     df.Race = df.Race.astype("category").cat.rename_categories({'CAUC': 'EA'})
     df.head()
```

```
[3]:          BrNum    RNum    Region  RIN    Age Sex Race       Dx  mitoRate  \
     R12864  Br1303  R12864  Caudate  9.6  42.98   F   AA  Schizo  0.032654
     R12865  Br1320  R12865  Caudate  9.5  53.12   M   AA  Schizo  0.019787
     R12866  Br1321  R12866  Caudate  9.1  57.13   F   AA  Schizo  0.013006
     R12867  Br1326  R12867  Caudate  9.2  74.56   M   AA  Schizo  0.032594
     R12868  Br1418  R12868  Caudate  9.5  43.35   M   AA  Schizo  0.052347

             rRNA_rate  overallMapRate
     R12864   0.000087        0.909350
     R12865   0.000070        0.873484
     R12866   0.000040        0.905505
     R12867   0.000038        0.910551
     R12868   0.000056        0.748659
```

```python
[4]: df.groupby("Region").size()
```

```
[4]: Region
     Caudate         420
     DLPFC           434
     DentateGyrus    161
     HIPPO           447
     dtype: int64
```

```
[5]: data = df.groupby(["Region", "Race"]).size().reset_index().rename(columns={0:
     ↪"N"})
     data
```

```
[5]:          Region Race    N
     0        Caudate   AA  217
     1        Caudate   EA  203
     2          DLPFC   AA  241
     3          DLPFC   EA  193
     4   DentateGyrus   AA   78
     5   DentateGyrus   EA   83
     6          HIPPO   AA  248
     7          HIPPO   EA  199
```

```
[6]: caudate = data[(data["Region"] == "Caudate")].drop("Region", axis=1).
     ↪set_index("Race")
     dlpfc = data[(data["Region"] == "DLPFC")].drop("Region", axis=1).
     ↪set_index("Race")
     gyrus = data[(data["Region"] == "DentateGyrus")].drop("Region", axis=1).
     ↪set_index("Race")
     hippocampus = data[(data["Region"] == "HIPPO")].drop("Region", axis=1).
     ↪set_index("Race")
```
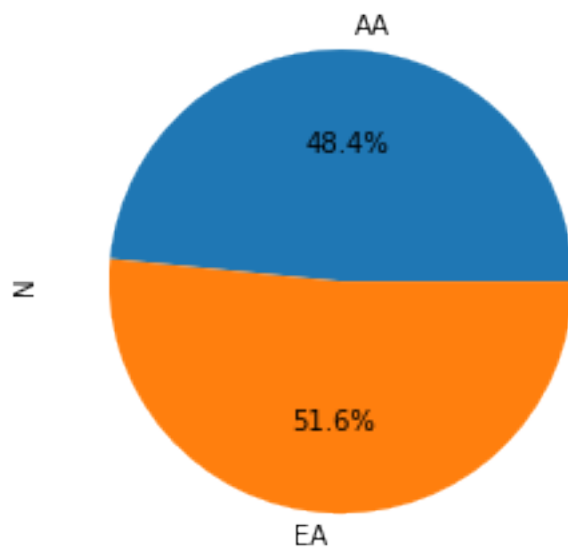
```
[7]: caudate.N.plot.pie(autopct="%.1f%%")
     plt.savefig('caudate_pie.pdf')
```
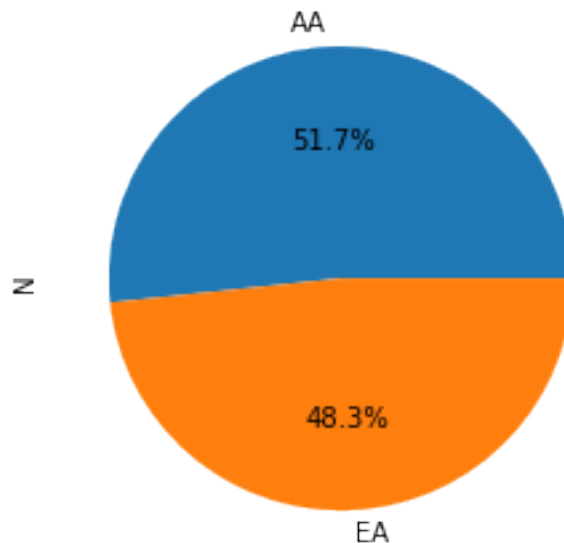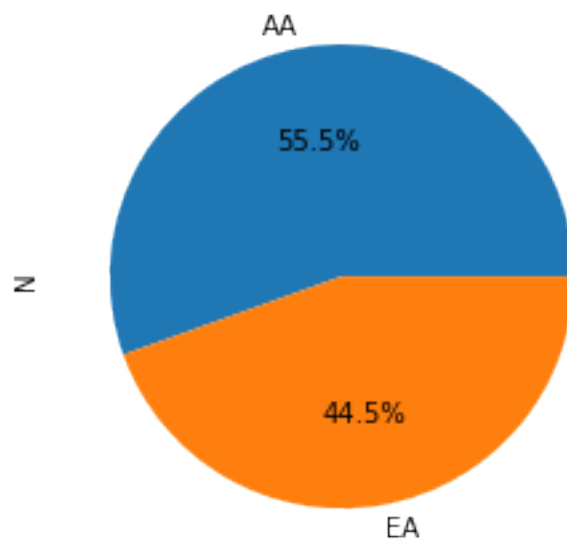
```
[8]: dlpfc.N.plot.pie(autopct="%.1f%%");
     plt.savefig('dlpfc_pie.pdf')
```

AA

55.5%

N

44.5%

EA
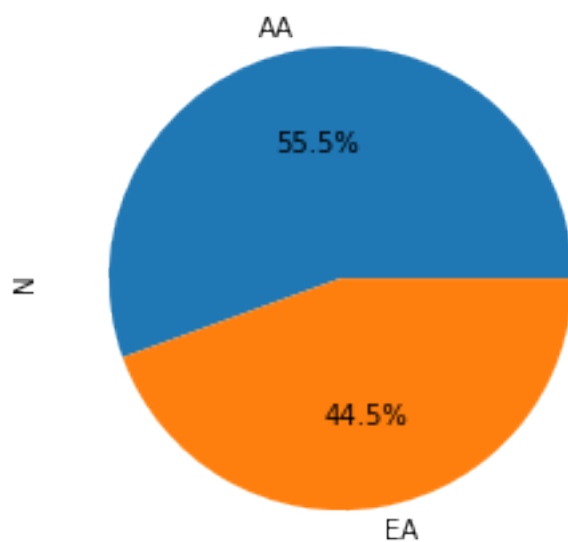
```
[9]: gyrus.N.plot.pie(autopct="%.1f%%");
     plt.savefig('dentate_gyrus_pie.pdf')
```

AA

48.4%

N

51.6%

EA

```
[10]: hippocampus.N.plot.pie(autopct="%.1f%%")
      plt.savefig('hippocampus_pie.pdf')
```



```
[ ]:
```