Making_RS_PCA

March 23, 2022

```
[1]: import pandas as pd
     import scipy.stats as stats
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sms
     import os
     import inspect
```

```
[2]: fname='../Data_In/ISTART_Combined_PersonalityData_020722.xlsx'
     df=pd.read_excel(fname)
     df = df[['participant_id', 'SPSRWD', 'BISBAS_BAS', 'SPSRWD.1', 'BISBAS_BAS.
      ↪1','BISBAS_BAS_Quintiles', 'SPSRWD_Quintiles', 'QUIN_DIFF']]
     bas_qcut=pd.to_numeric(pd.cut(
         df['BISBAS_BAS.1'],bins=[0,34,37,39,42,100],
         labels=[1,2,3,4,5],include_lowest=True))

     spsq_qcut=pd.to_numeric(pd.cut(
         df['SPSRWD.1'],bins=[0,8,11,13.6845, 16.6957,100],
         labels=[1,2,3,4,5],include_lowest=True))

     df['BISBAS_BAS_Q.1']=bas_qcut
     df['SPSRWD_Q.1']=spsq_qcut
     df['QuinDiff.1']=df['BISBAS_BAS_Q.1']-df['SPSRWD_Q.1']

     #Remove counter rows
     p_idToremove=['N Missing', 'N per column']
     df.replace(to_replace=999, value=np.nan,inplace=True)
     df = df[df.participant_id.isin(p_idToremove) == False]
     df.head()
```

```
[2]:    participant_id  SPSRWD  BISBAS_BAS  SPSRWD.1  BISBAS_BAS.1  \
     0            1001    41.0        44.0       7.0          44.0
     1            1002    33.0        35.0      15.0          35.0
     2            1003    42.0        45.0       6.0          45.0
     3            1004    14.0        39.0      14.0          39.0
     4            1006    12.0        37.0      12.0          37.0
```

```
      BISBAS_BAS_Quintiles  SPSRWD_Quintiles  QUIN_DIFF  BISBAS_BAS_Q.1  \
0                        5               5.0        0.0               5
1                        2               2.0        0.0               2
2                        5               5.0        0.0               5
3                        4               4.0        0.0               3
4                        3               3.0        0.0               2

      SPSRWD_Q.1  QuinDiff.1
0              1           4
1              4          -2
2              1           4
3              4          -1
4              3          -1
```

[3]:
```
cormat=df.corr()
display(cormat)
fig = plt.figure(figsize=(12, 8))
sms.heatmap(cormat,cmap='vlag',vmin=-1)
```

```
                        SPSRWD  BISBAS_BAS  SPSRWD.1  BISBAS_BAS.1  \
SPSRWD                1.000000    0.074672 -0.148400     -0.209165
BISBAS_BAS            0.074672    1.000000  0.634012      0.917155
SPSRWD.1             -0.148400    0.634012  1.000000      0.676969
BISBAS_BAS.1         -0.209165    0.917155  0.676969      1.000000
BISBAS_BAS_Quintiles  0.067583    0.906164  0.637191      0.814639
SPSRWD_Quintiles      0.554094    0.807275  0.747241      0.807275
QUIN_DIFF            -0.092702    0.352355 -0.014545      0.352355
BISBAS_BAS_Q.1       -0.185785    0.862328  0.634014      0.908738
SPSRWD_Q.1           -0.217998    0.599824  0.957904      0.663307
QuinDiff.1            0.028178    0.343364 -0.342387      0.325938

                        BISBAS_BAS_Quintiles  SPSRWD_Quintiles  QUIN_DIFF  \
SPSRWD                              0.067583          0.554094  -0.092702
BISBAS_BAS                          0.906164          0.807275   0.352355
SPSRWD.1                            0.637191          0.747241  -0.014545
BISBAS_BAS.1                        0.814639          0.807275   0.352355
BISBAS_BAS_Quintiles                1.000000          0.880203   0.410917
SPSRWD_Quintiles                    0.880203          1.000000  -0.070988
QUIN_DIFF                           0.410917         -0.070988   1.000000
BISBAS_BAS_Q.1                      0.864853          0.821482   0.355463
SPSRWD_Q.1                          0.624527          0.728440   0.020916
QuinDiff.1                          0.317963          0.147403   0.392489

                        BISBAS_BAS_Q.1  SPSRWD_Q.1  QuinDiff.1
SPSRWD                       -0.185785   -0.217998    0.028178
BISBAS_BAS                    0.862328    0.599824    0.343364
SPSRWD.1                      0.634014    0.957904   -0.342387
```
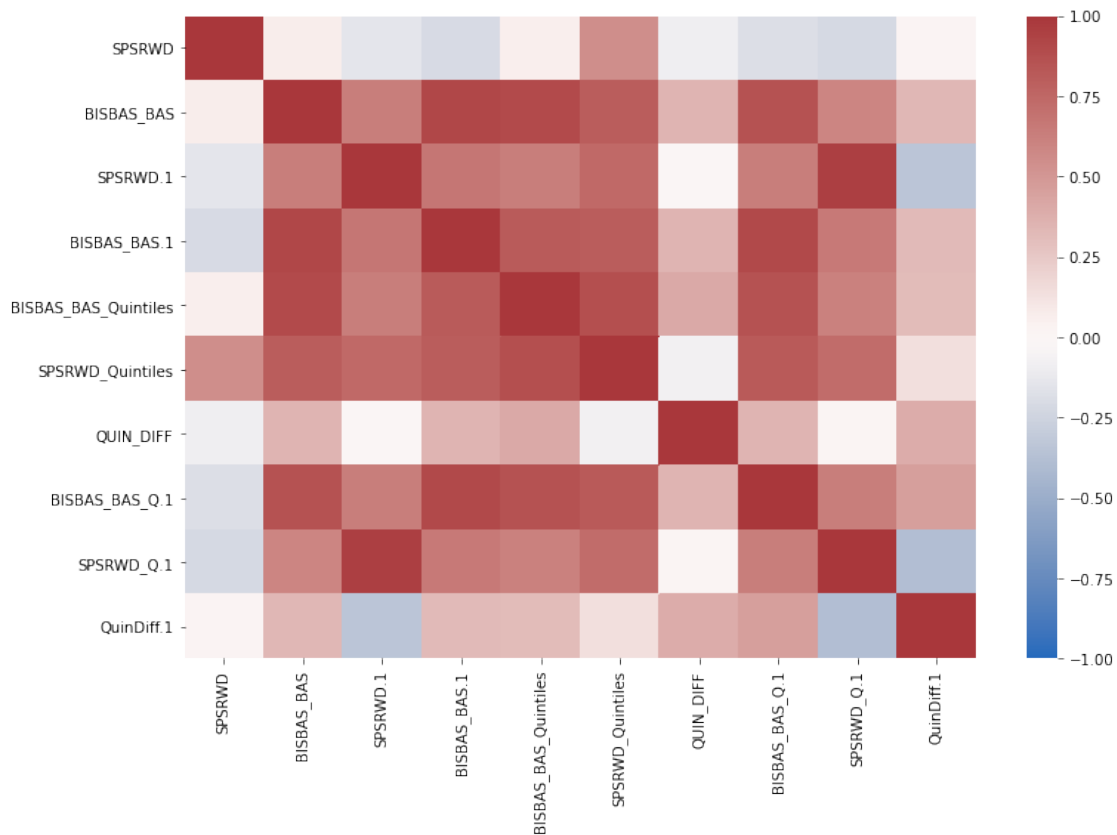
```
BISBAS_BAS.1            0.908738    0.663307    0.325938
BISBAS_BAS_Quintiles   0.864853    0.624527    0.317963
SPSRWD_Quintiles       0.821482    0.728440    0.147403
QUIN_DIFF              0.355463    0.020916    0.392489
BISBAS_BAS_Q.1         1.000000    0.638362    0.464038
SPSRWD_Q.1            0.638362    1.000000   -0.385620
QuinDiff.1            0.464038   -0.385620    1.000000
```
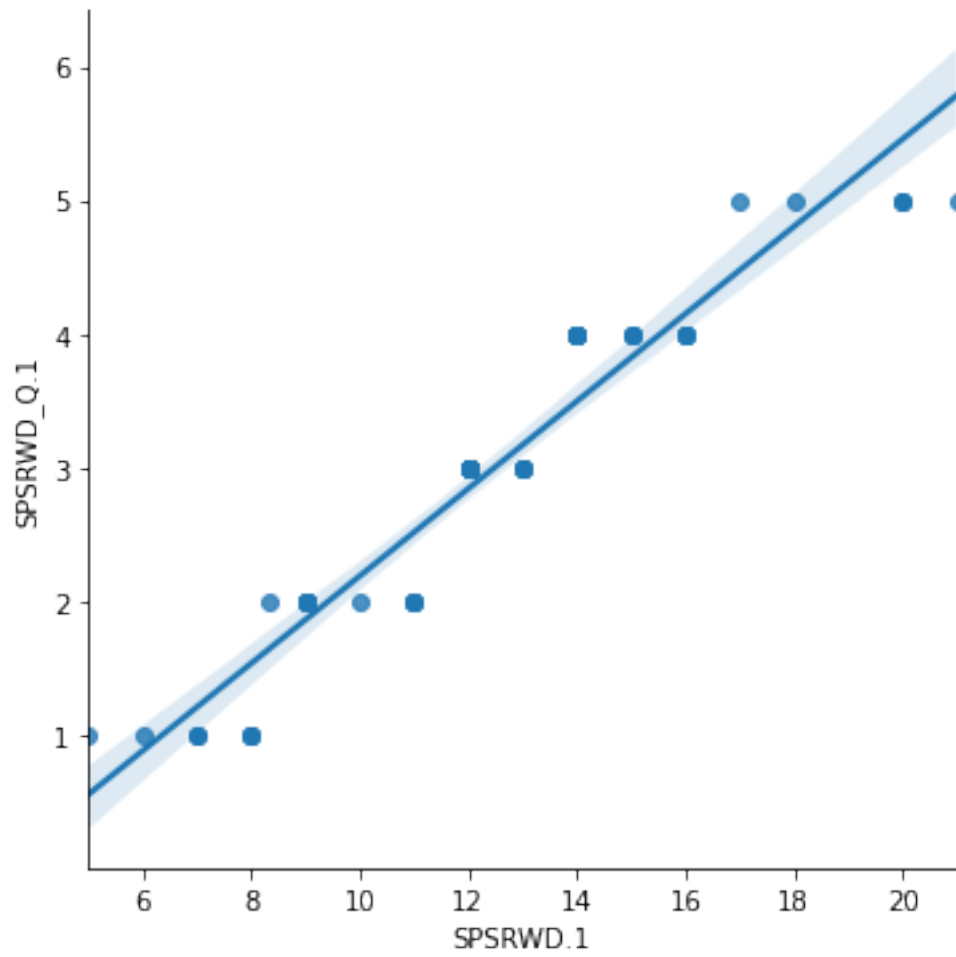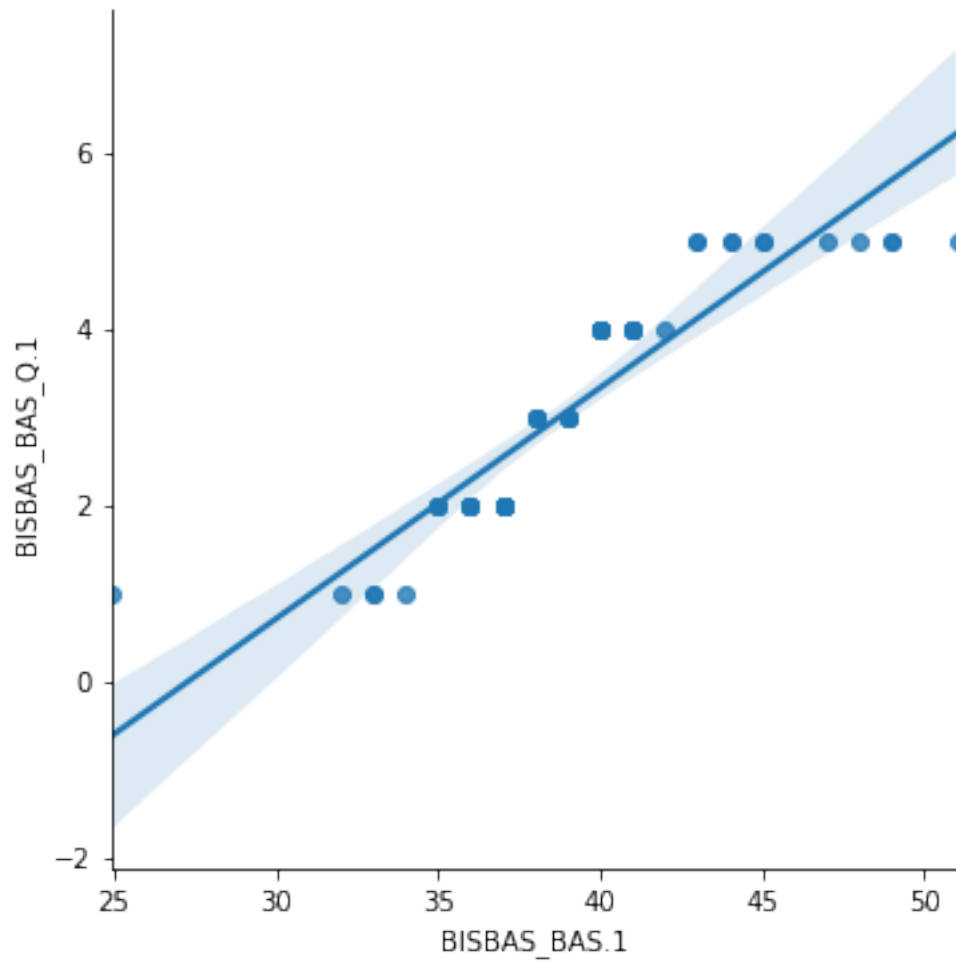
[3]: <AxesSubplot:>


```
[4]: g=sms.lmplot(x='SPSRWD.1',y='SPSRWD_Quintiles',data=df)
     plt.show()
     g=sms.lmplot(x='BISBAS_BAS.1',y='BISBAS_BAS_Quintiles',data=df)
     plt.show()
     g=sms.lmplot(x='SPSRWD.1',y='SPSRWD_Q.1',data=df)
     plt.show()
     g=sms.lmplot(x='BISBAS_BAS.1',y='BISBAS_BAS_Q.1',data=df)
     plt.show()
```
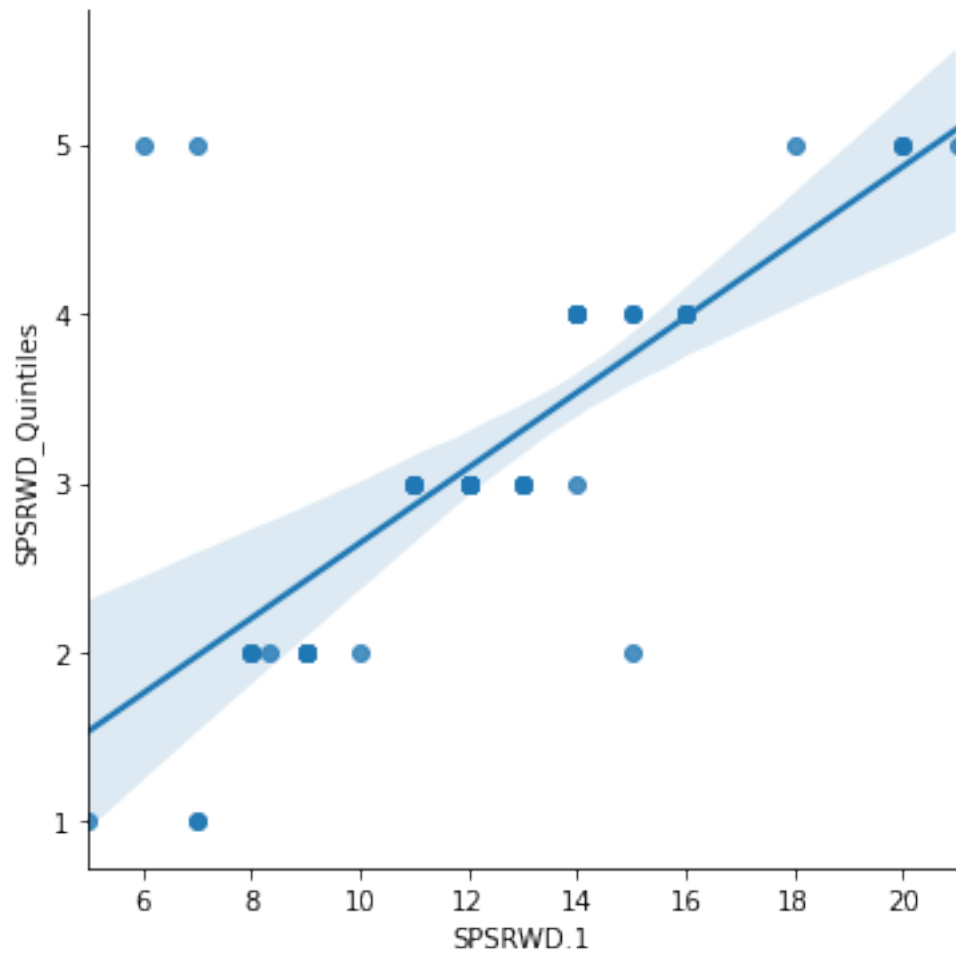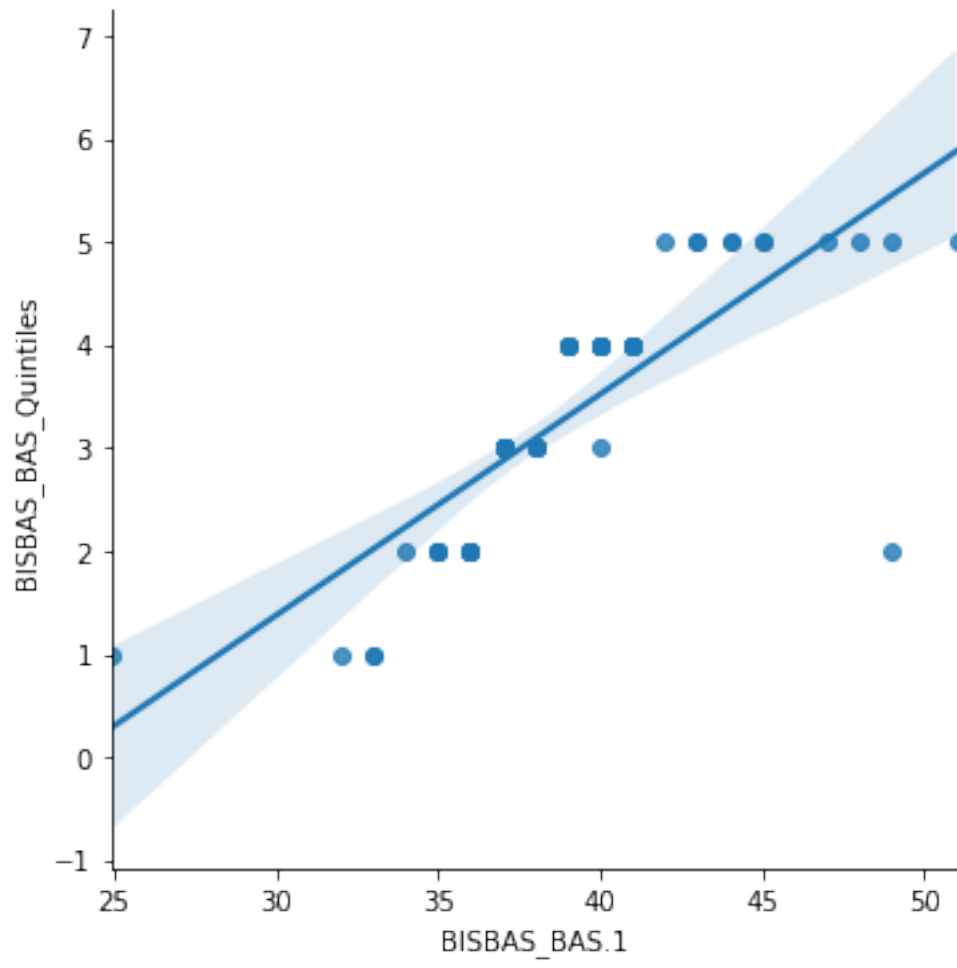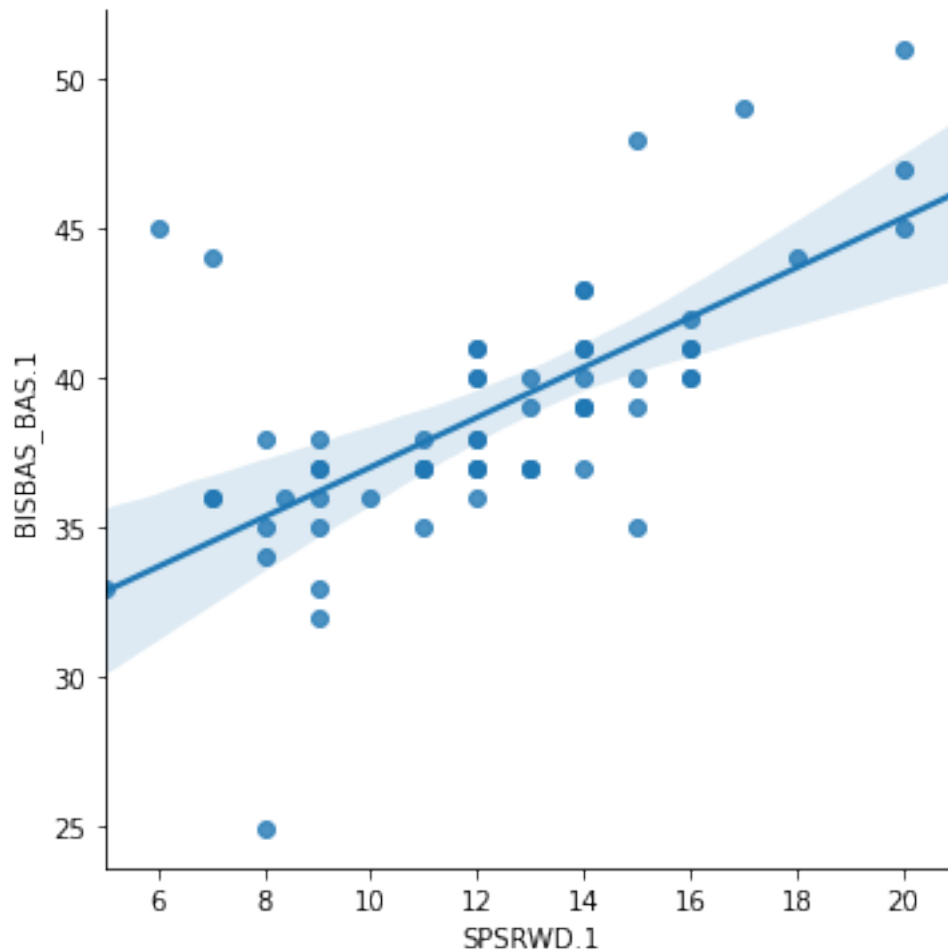
```
[5]: r,p=stats.pearsonr(df['SPSRWD.1'],df['BISBAS_BAS.1'])
     print("Correlation between reward sensitivity is \n R^2: %s p: %s"%(r**2,p))
     g=sms.lmplot(x='SPSRWD.1',y='BISBAS_BAS.1',data=df)
     plt.show()
     print(type(g))
```

Correlation between reward sensitivity is
 R^2: 0.4582872755754281 p: 1.5333008566420115e-09

```
<class 'seaborn.axisgrid.FacetGrid'>
```

0.1 composite raw included

```python
[6]: from sklearn.decomposition import PCA
     from scipy.stats import zscore
     pca=PCA(n_components=1)

     features=['SPSRWD.1','BISBAS_BAS.1']
     X = df.loc[:, features].values
     PC1=pca.fit_transform(X)

     df['PC1_raw_included']=PC1
     plt.scatter(X[:, 0], X[:, 1], alpha=0.3, label="samples")
     for i, (comp, var) in enumerate(zip(pca.components_, pca.explained_variance_)):
         comp = comp * var  # scale component by its variance explanation power
         plt.plot(
```
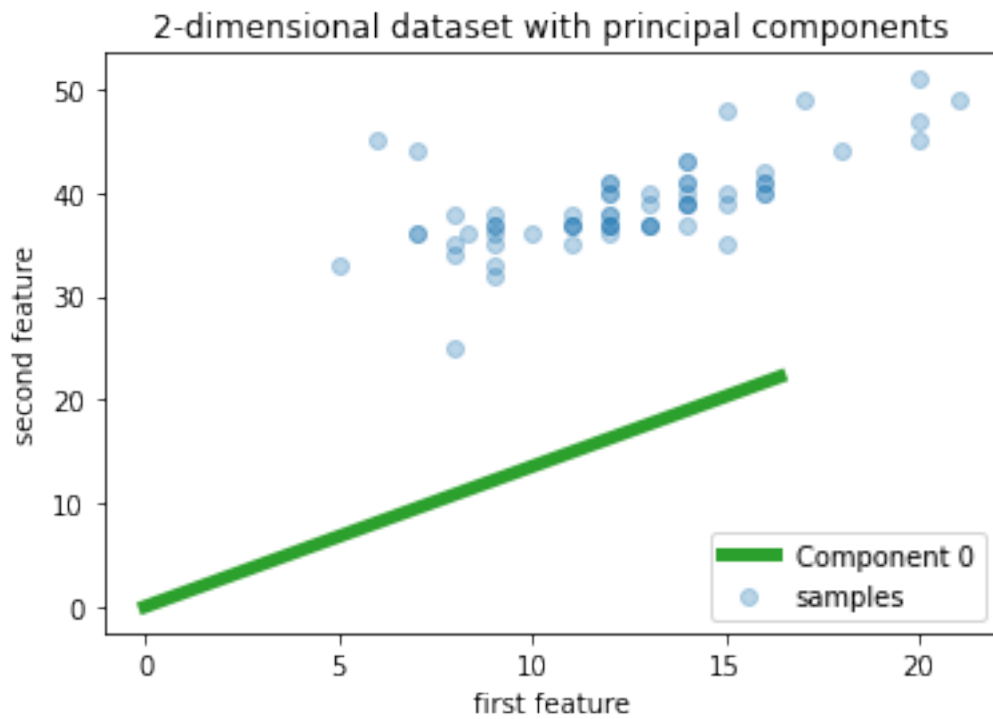
```
        [0, comp[0]],
        [0, comp[1]],
        label=f"Component {i}",
        linewidth=5,
        color=f"C{i + 2}",
    )
plt.gca().set(
    title="2-dimensional dataset with principal components",
    xlabel="first feature",
    ylabel="second feature",
)
plt.legend()
plt.show()
```


0.2 composite zscore included

```
[7]: features=['SPSRWD.1','BISBAS_BAS.1']
     X = df.loc[:, features].apply(zscore).values
     PC1=pca.fit_transform(X)


     df['PC1_score_included']=PC1
```

```
plt.scatter(X[:, 0], X[:, 1], alpha=0.3, label="samples")
for i, (comp, var) in enumerate(zip(pca.components_, pca.explained_variance_)):
    comp = comp * var  # scale component by its variance explanation power
    plt.plot(
        [0, comp[0]],
        [0, comp[1]],
        label=f"Component {i}",
        linewidth=5,
        color=f"C{i + 2}",
    )
plt.gca().set(
    title="2-dimensional dataset with principal components",
    xlabel="first feature",
    ylabel="second feature",
)
plt.legend()
plt.show()
```
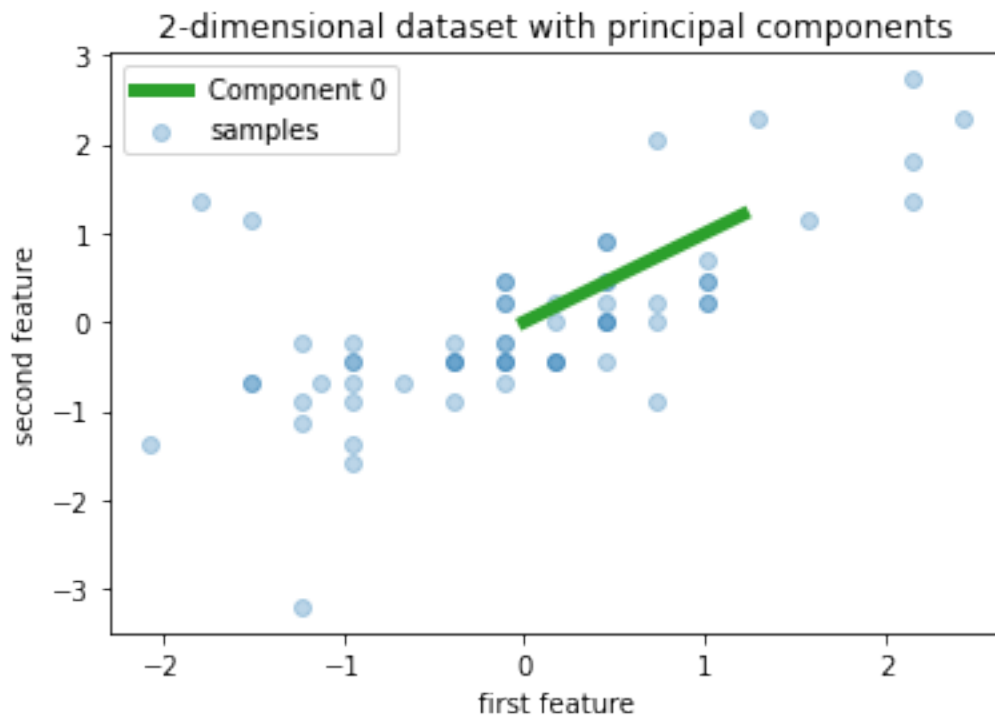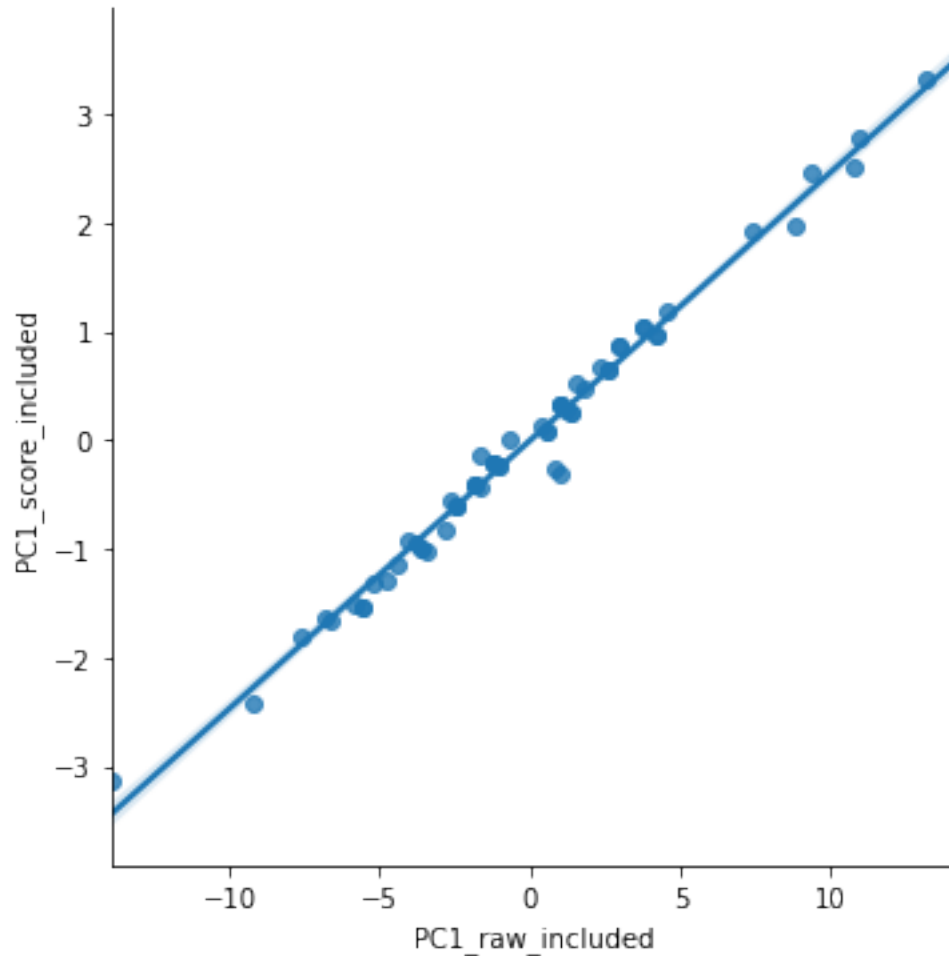


```
[8]: df.head()
     r,p=stats.pearsonr(df['PC1_raw_included'],df['PC1_score_included'])
     print("Correlation between reward sensitivity is \n R^2: %s p: %s"%(r**2,p))
     g=sms.lmplot(x='PC1_raw_included',y='PC1_score_included',data=df)
     plt.show()
```

```
print(type(g))
```

Correlation between reward sensitivity is
 R^2: 0.9880895899979261 p: 1.9560220980791162e-59


```
<class 'seaborn.axisgrid.FacetGrid'>
```

0.3   composite raw excluded

```
[15]: ex_df=df[np.abs(df['QuinDiff.1'])<=1]

      features=['SPSRWD.1','BISBAS_BAS.1']
      X = ex_df.loc[:, features].values
      pca.fit(X)

      X = df.loc[:, features].values
      PC1=pca.transform(X)
```
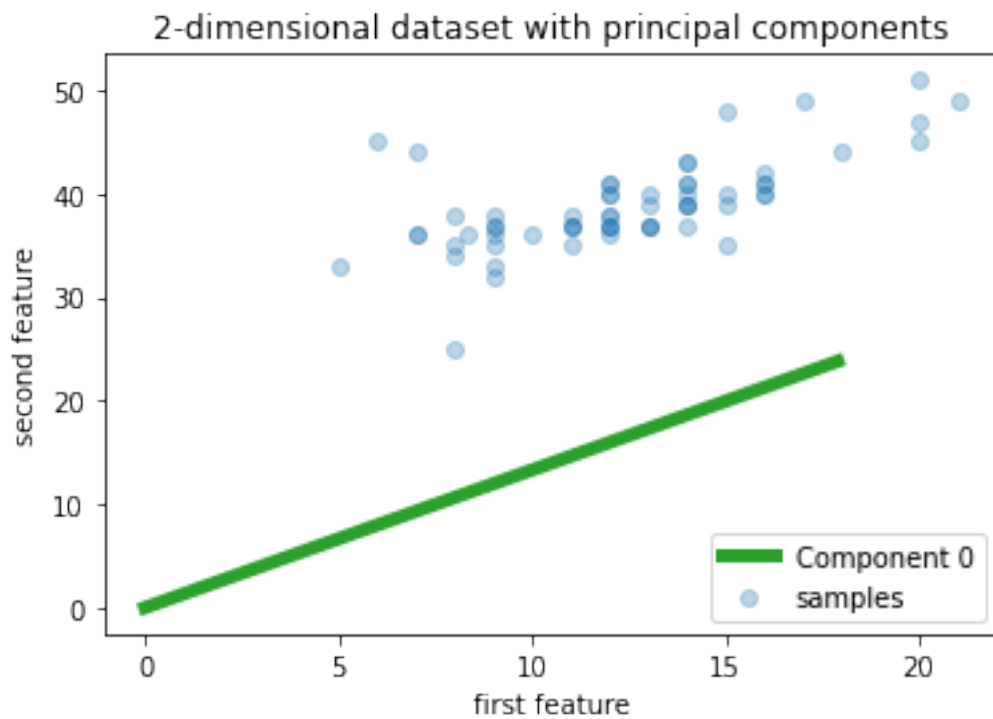
```
df['PC1_raw_excluded']=PC1
plt.scatter(X[:, 0], X[:, 1], alpha=0.3, label="samples")
for i, (comp, var) in enumerate(zip(pca.components_, pca.explained_variance_)):
    comp = comp * var  # scale component by its variance explanation power
    plt.plot(
        [0, comp[0]],
        [0, comp[1]],
        label=f"Component {i}",
        linewidth=5,
        color=f"C{i + 2}",
    )
plt.gca().set(
    title="2-dimensional dataset with principal components",
    xlabel="first feature",
    ylabel="second feature",
)
plt.legend()
plt.show()
```


2-dimensional dataset with principal components

0.4 composite zscore excluded

```
[17]: ex_df=df[np.abs(df['QuinDiff.1'])<=1]

      features=['SPSRWD.1','BISBAS_BAS.1']
      X = ex_df.loc[:, features].apply(zscore).values
      pca.fit(X)

      X = df.loc[:, features].apply(zscore).values
      PC1=pca.transform(X)
      df['PC1_score_excluded']=PC1
      plt.scatter(X[:, 0], X[:, 1], alpha=0.3, label="samples")
      for i, (comp, var) in enumerate(zip(pca.components_, pca.explained_variance_)):
          comp = comp * var  # scale component by its variance explanation power
          plt.plot(
              [0, comp[0]],
              [0, comp[1]],
              label=f"Component {i}",
              linewidth=5,
              color=f"C{i + 2}",
          )
      plt.gca().set(
          title="2-dimensional dataset with principal components",
          xlabel="first feature",
          ylabel="second feature",
      )
      plt.legend()
      plt.show()
```
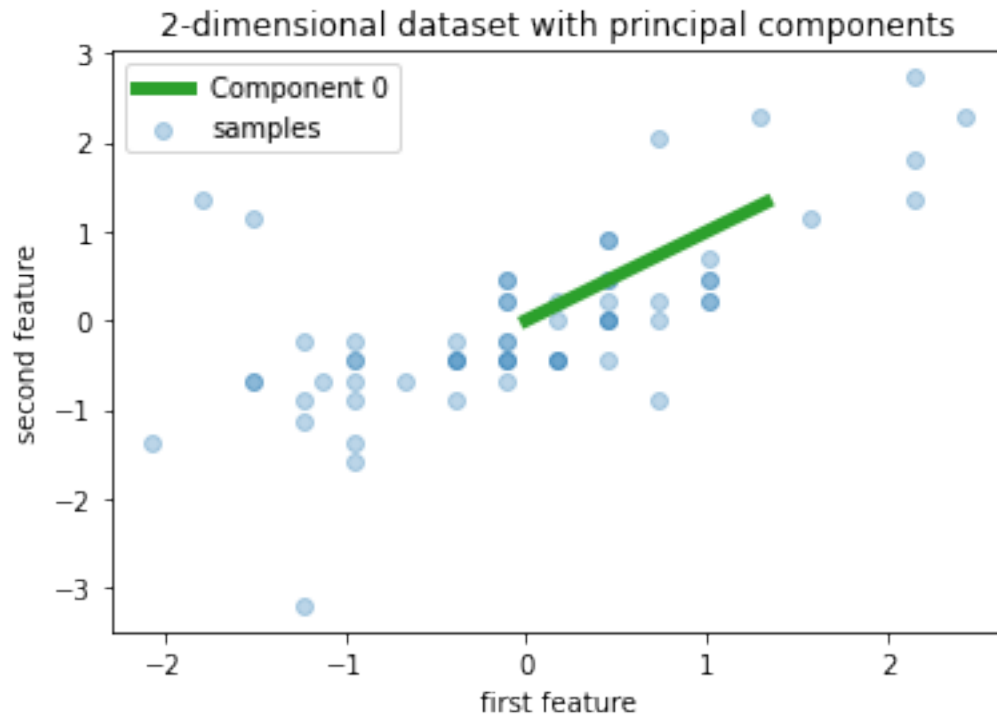
2-dimensional dataset with principal components

```
[22]: df.head()
      r,p=stats.pearsonr(df['PC1_raw_excluded'],df['PC1_score_excluded'])
      print("Correlation between reward sensitivity is \n R^2: %s p: %s"%(r**2,p))
      g=sms.lmplot(x='PC1_raw_excluded',y='PC1_score_excluded',data=df)
      plt.show()
      print(type(g))
```

Correlation between reward sensitivity is
 R^2: 0.9888853289926454 p: 2.456414553170082e-60

```
<class 'seaborn.axisgrid.FacetGrid'>
```
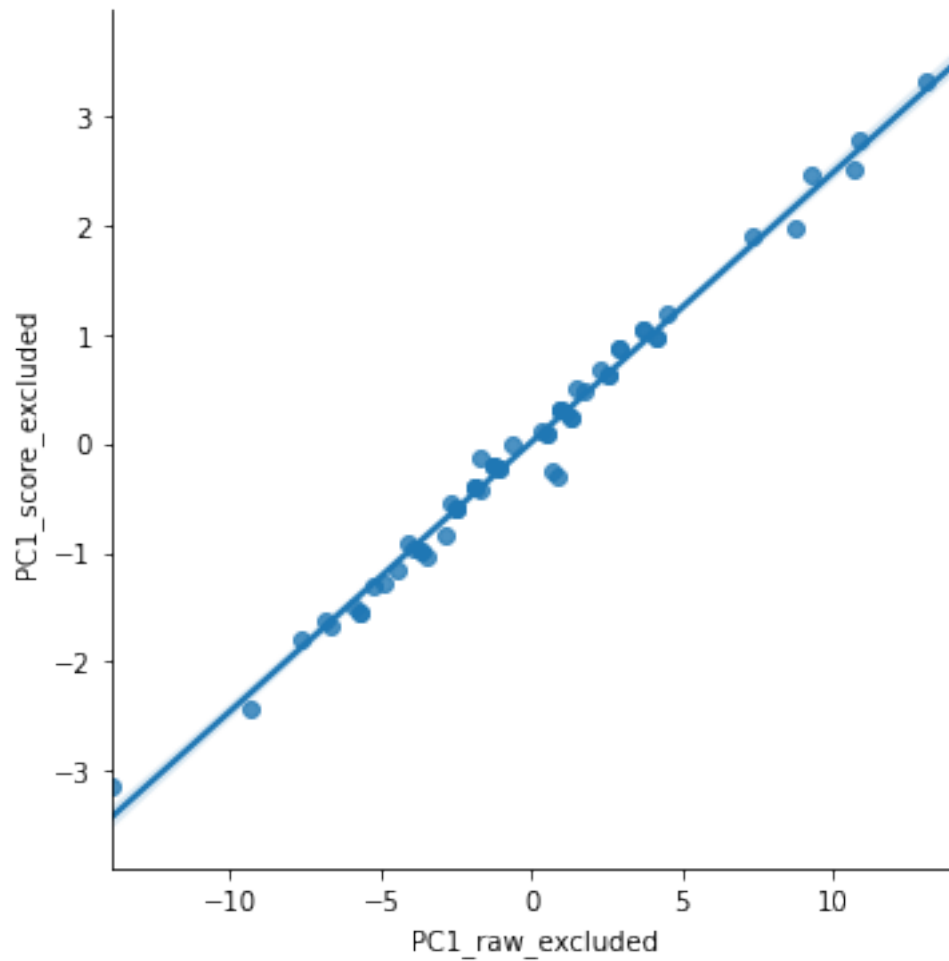
[21]:
```python
df.head()
r,p=stats.pearsonr(df['PC1_score_included'],df['PC1_score_excluded'])
print("Correlation between reward sensitivity is \n R^2: %s p: %s"%(r**2,p))
g=sms.lmplot(x='PC1__included',y='PC1_score_included',data=df)
plt.show()
print(type(g))
```
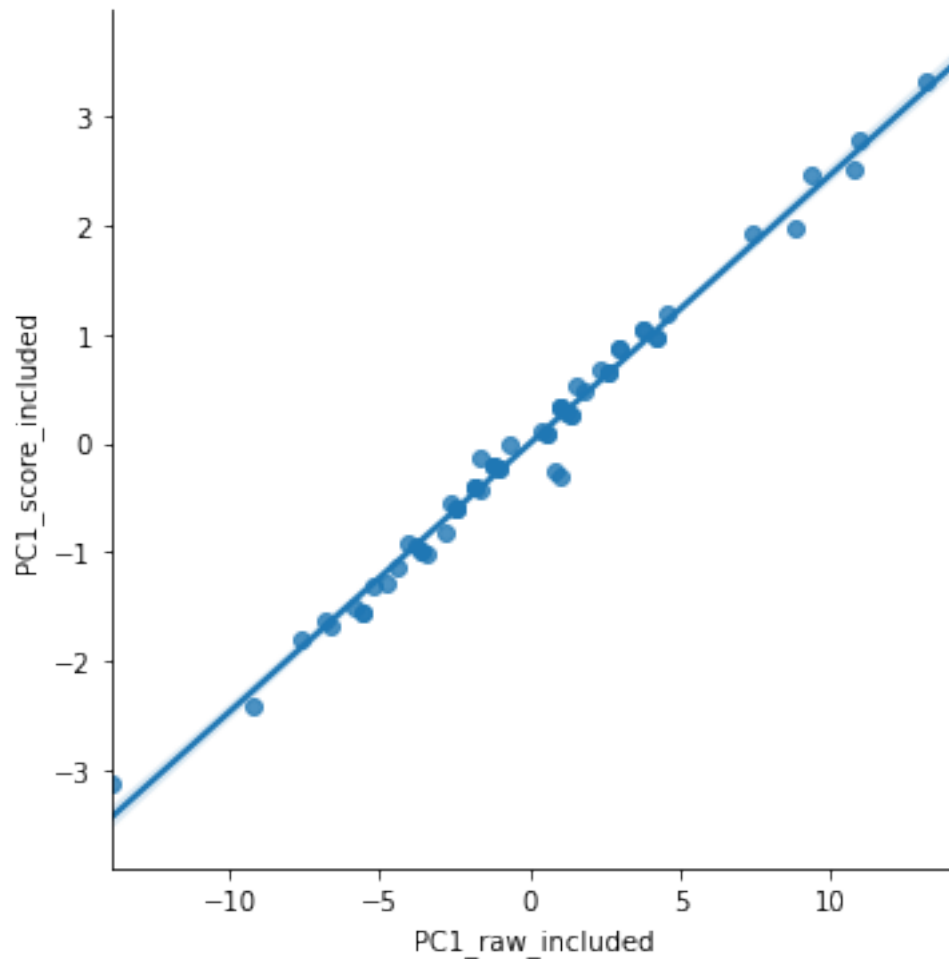
```
Correlation between reward sensitivity is
 R^2: 1.0 p: 0.0
```

```
[19]: cormat=df.corr()
      display(cormat)
      fig = plt.figure(figsize=(12, 8))
      sms.heatmap(cormat,cmap='vlag',vmin=-1)
```

	SPSRWD	BISBAS_BAS	SPSRWD.1	BISBAS_BAS.1 \
SPSRWD	1.000000	0.074672	-0.148400	-0.209165
BISBAS_BAS	0.074672	1.000000	0.634012	0.917155
SPSRWD.1	-0.148400	0.634012	1.000000	0.676969
BISBAS_BAS.1	-0.209165	0.917155	0.676969	1.000000
BISBAS_BAS_Quintiles	0.067583	0.906164	0.637191	0.814639
SPSRWD_Quintiles	0.554094	0.807275	0.747241	0.807275
QUIN_DIFF	-0.092702	0.352355	-0.014545	0.352355
BISBAS_BAS_Q.1	-0.185785	0.862328	0.634014	0.908738
SPSRWD_Q.1	-0.217998	0.599824	0.957904	0.663307

```
QuinDiff.1            0.028178    0.343364 -0.342387     0.325938
PC1_raw_included     -0.202328    0.880381  0.866358     0.954079
PC1_score_included   -0.195244    0.846995  0.915688     0.915688
PC1_raw_excluded     -0.202126    0.879413  0.868215     0.952955
PC1_score_excluded   -0.195244    0.846995  0.915688     0.915688


                      BISBAS_BAS_Quintiles  SPSRWD_Quintiles  QUIN_DIFF  \
SPSRWD                            0.067583          0.554094  -0.092702
BISBAS_BAS                       0.906164          0.807275   0.352355
SPSRWD.1                         0.637191          0.747241  -0.014545
BISBAS_BAS.1                     0.814639          0.807275   0.352355
BISBAS_BAS_Quintiles             1.000000          0.880203   0.410917
SPSRWD_Quintiles                 0.880203          1.000000  -0.070988
QUIN_DIFF                        0.410917         -0.070988   1.000000
BISBAS_BAS_Q.1                   0.864853          0.821482   0.355463
SPSRWD_Q.1                       0.624527          0.728440   0.020916
QuinDiff.1                       0.317963          0.147403   0.392489
PC1_raw_included                 0.812112          0.854634   0.231302
PC1_score_included               0.792753          0.851138   0.182136
PC1_raw_excluded                 0.811610          0.854692   0.229663
PC1_score_excluded               0.792753          0.851138   0.182136


                     BISBAS_BAS_Q.1  SPSRWD_Q.1  QuinDiff.1  \
SPSRWD                    -0.185785   -0.217998    0.028178
BISBAS_BAS                 0.862328    0.599824    0.343364
SPSRWD.1                   0.634014    0.957904   -0.342387
BISBAS_BAS.1               0.908738    0.663307    0.325938
BISBAS_BAS_Quintiles       0.864853    0.624527    0.317963
SPSRWD_Quintiles           0.821482    0.728440    0.147403
QUIN_DIFF                  0.355463    0.020916    0.392489
BISBAS_BAS_Q.1             1.000000    0.638362    0.464038
SPSRWD_Q.1                 0.638362    1.000000   -0.385620
QuinDiff.1                 0.464038   -0.385620    1.000000
PC1_raw_included           0.874670    0.839955    0.081815
PC1_score_included         0.842400    0.885242   -0.008982
PC1_raw_excluded           0.873739    0.841668    0.078727
PC1_score_excluded         0.842400    0.885242   -0.008982


                     PC1_raw_included  PC1_score_included  PC1_raw_excluded  \
SPSRWD                       -0.202328           -0.195244         -0.202126
BISBAS_BAS                    0.880381            0.846995          0.879413
SPSRWD.1                      0.866358            0.915688          0.868215
BISBAS_BAS.1                  0.954079            0.915688          0.952955
BISBAS_BAS_Quintiles          0.812112            0.792753          0.811610
SPSRWD_Quintiles              0.854634            0.851138          0.854692
QUIN_DIFF                     0.231302            0.182136          0.229663
BISBAS_BAS_Q.1                0.874670            0.842400          0.873739
SPSRWD_Q.1                    0.839955            0.885242          0.841668
```
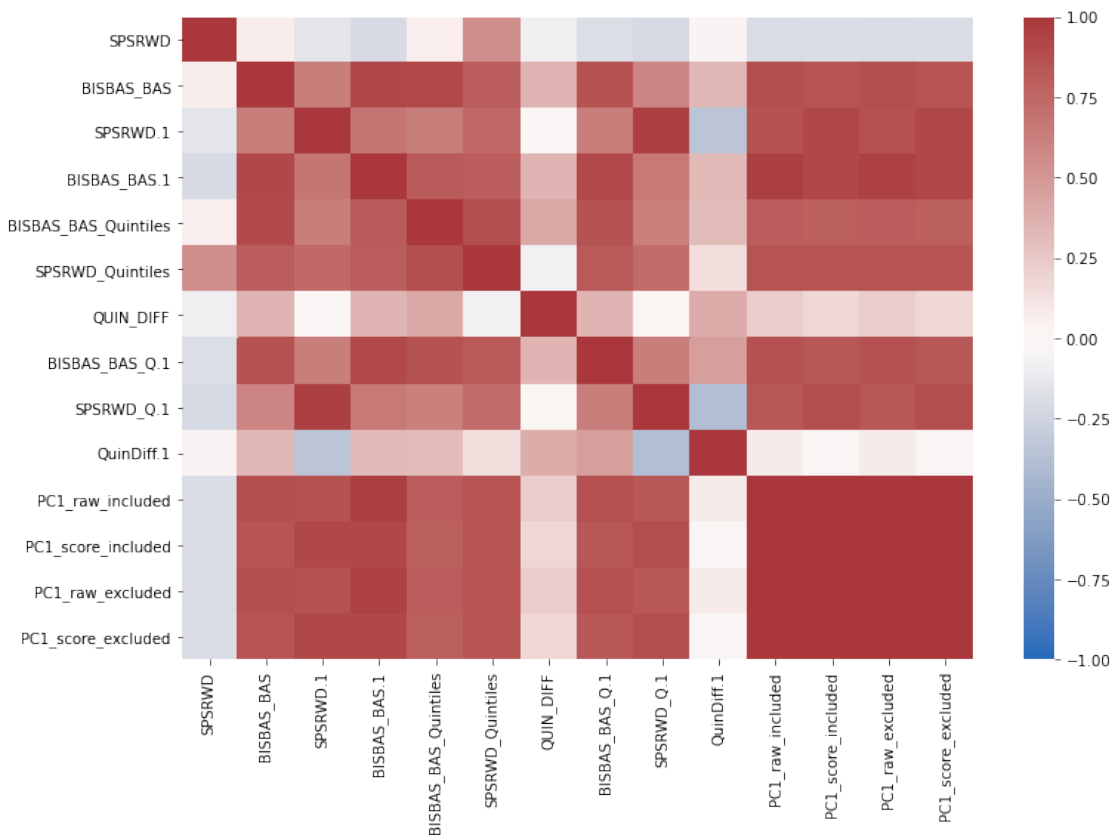
```
QuinDiff.1                         0.081815            -0.008982             0.078727
PC1_raw_included                   1.000000             0.994027             0.999993
PC1_score_included                 0.994027             1.000000             0.994427
PC1_raw_excluded                   0.999993             0.994427             1.000000
PC1_score_excluded                 0.994027             1.000000             0.994427


                         PC1_score_excluded
SPSRWD                            -0.195244
BISBAS_BAS                         0.846995
SPSRWD.1                           0.915688
BISBAS_BAS.1                       0.915688
BISBAS_BAS_Quintiles               0.792753
SPSRWD_Quintiles                   0.851138
QUIN_DIFF                          0.182136
BISBAS_BAS_Q.1                     0.842400
SPSRWD_Q.1                         0.885242
QuinDiff.1                        -0.008982
PC1_raw_included                   0.994027
PC1_score_included                 1.000000
PC1_raw_excluded                   0.994427
PC1_score_excluded                 1.000000
```

[19]: <AxesSubplot:>

[ ]: