

day9

December 25, 2022

```
[1]: import pandas as pd

[2]: data = pd.read_csv("human_phosphorylations_08_12_22.tsv", sep="\t")

[3]: d= data.loc[data['TYPEA'] == "protein"] #pull out columns labelled as protein
      ↪from TYPE A column

[4]: dff= d.loc[d['MECHANISM'] == "phosphorylation"] #pull out labels called
      ↪phosphorylation in mechanism

[5]: df= dff.loc[dff['TYPEB'] == "protein"] # pull out columns labelled protein from
      ↪TYPE B

[6]: DF = df[['ENTITYA', 'MECHANISM', 'ENTITYB', 'RESIDUE']]

[7]: DF= DF.rename(columns={'ENTITYA': 'Kinase', 'MECHANISM': 'Action', 'ENTITYB':
      ↪'Target', 'RESIDUE': 'Residue'})

[8]: DF.to_csv("originaldata.tsv", sep="\t", index= False)

[9]: DF.to_csv("originaldata.sif", sep="\t", index= None, header=False)

[10]: Kinase= DF['Kinase']

[11]: DF2 = DF.loc[DF["Target"].isin(Kinase)] #pulls out only what is in target which
      ↪is ALSO in kinase

[12]: DF2.to_csv("kinasematched.tsv", sep="\t", index= False)

[13]: DF2.to_csv("kinasematched.sif", sep="\t", index= None, header=False) #kinase
      ↪matched should make data look less cluttered

[14]: #TASK 4

[15]: nomapk= DF2.loc[DF2['Kinase'] != "MAPK1"] #taking mapk1 out of kinase column

[16]: nomapk.to_csv("nomapk.sif", sep="\t", index= None, header=False)
```

```
[17]: mapk= DF2.loc[DF2['Kinase'] == 'MAPK1'] #dataframe of all the MAPK1 connections_
      ↪we removed above
```

```
[18]: mapk.to_csv("mapk.sif", sep="\t", index= None, header=False)
```

```
[19]: #TASK 4A
```

```
[20]: ser642_network= nomapk.loc[nomapk['Residue'] == "Ser642"] #pull out Ser642 WEE1
      ser642_network
```

```
[20]:      Kinase      Action Target Residue
      8507  BRSK1 phosphorylation  WEE1 Ser642
```

```
[21]: ser642_network.to_csv("ser642.sif", sep="\t", index= None, header=False)
```

```
[22]: ser642_network.iloc[0,3] = 'Gln642' #change residue
      ser642_network
```

```
/tmp/ipykernel_305/3467901808.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
      ser642_network.iloc[0,3] = 'Gln642' #change residue
```

```
[22]:      Kinase      Action Target Residue
      8507  BRSK1 phosphorylation  WEE1 Gln642
```

```
[23]: gln642_network= ser642_network
```

```
[24]: gln642_network.to_csv("gln642.sif", sep="\t", index= None, header=False)
```

```
[26]: #TASK 4B
```

```
[27]: tyr474= DF.loc[DF['Target'] == 'AKT1']
```

```
[28]: tyr474.loc[tyr474['Residue'] == 'Tyr474'] #no residue Tyr474 for AKT1
```

```
[28]: Empty DataFrame
      Columns: [Kinase, Action, Target, Residue]
      Index: []
```

```
[29]: degree_data= pd.read_csv('degreedata.csv')
```

```
[30]: degree_data['Degree'].plot(kind='hist', logy=True, rwidth=0.9, x= 'Degree')
```

```
[30]: <AxesSubplot: ylabel='Frequency'>
```

