

Riley Payung, Salem Aluswaidi

Professor Jason Kinser

CDS 411

October 25, 2020

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import floyd
import bacon
import pandas as pd
import movies3 as mvs
```

```
In [2]: movies, actors, isin = mvs.ReadData("movies1000.xlsx")
```

```
In [91]: def moviesByYear(year,movies):
    moviesResult = []
    for i in movies:
        if (i[2]==year):
            moviesResult.append(i[0]);
    return moviesResult;
```

```
In [80]: movies2007 = moviesByYear(2007,movies)
```

```
In [82]: aids = mvs.AidsFromMids(isin,movies2007)
```

```
In [8]: def MakeG(movies, isin):
    mat = np.array(movies);
    t = mat[:,0]+0
    mids = np.array(list(set(t)))
    N = len(mids)
    G = np.zeros((N+1,N+1))

    for i in mids:
        aids = mvs.AidsFromMid(isin, i);
        mids2 = mvs.MidsFromAids(isin, aids);
        for j in mids2:
            G[i-1,j-1] = 1;
    return G;
```

```
In [9]: G = MakeG(movies,isin)
```

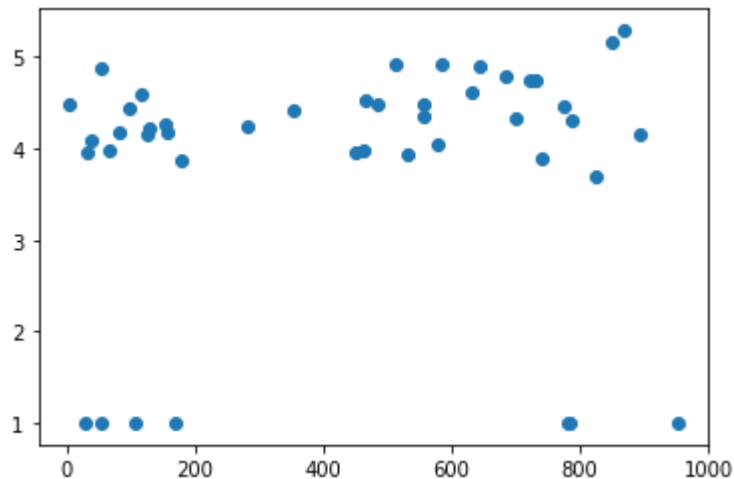
```
In [10]: g,p = bacon.RunFloyd(G)
```

```
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950
1000
```

```
In [48]: def YACL(mids):
AverageChainLength = [];
count = 0;
pathlength = 0;
for j in mids:
    for i in movies:
        pathlength += len(floyd.FindPath(p,j,i[0]))
        count += 1;
    AverageChainLength.append(pathlength/count);
    pathlength = 0;
    count = 0;
return AverageChainLength;
```

```
In [83]: yac1 = YACL(movies2007)
```

```
In [84]: plt.scatter(movies2007,yac1)
plt.show()
```



```
In [85]: np.mean(yac1) # 2007's YACL
```

```
Out[85]: 3.8526
```

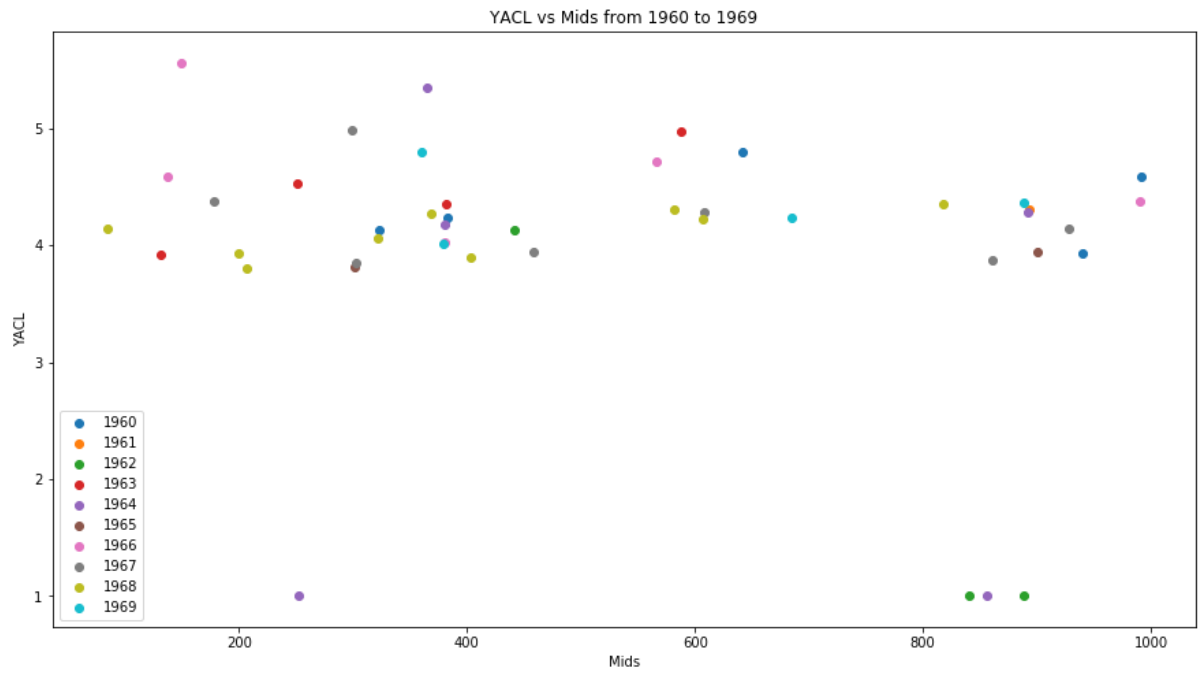
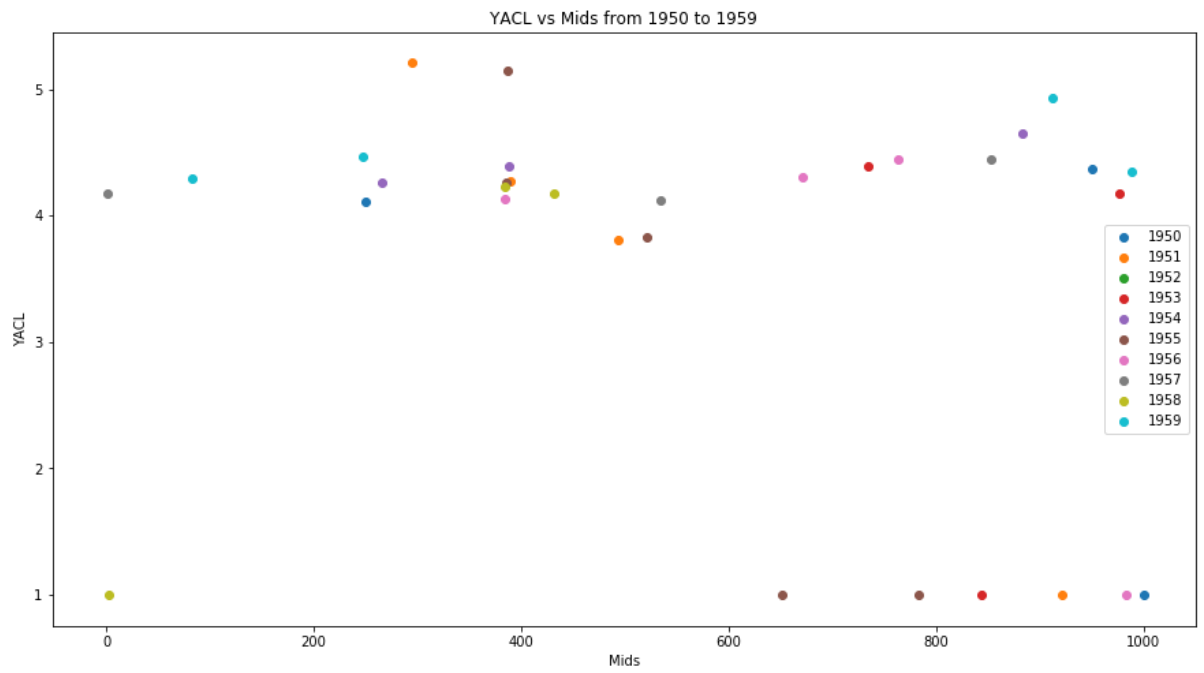
```
In [102]: def Run(yearLow, yearHigh):
midsYears = [];
yac1Years = [];
for i in range(yearLow,yearHigh):
    mids = moviesByYear(i,movies);
    midsYears.append(mids);
    yac1Years.append(YACL(mids));
return midsYears, yac1Years
```

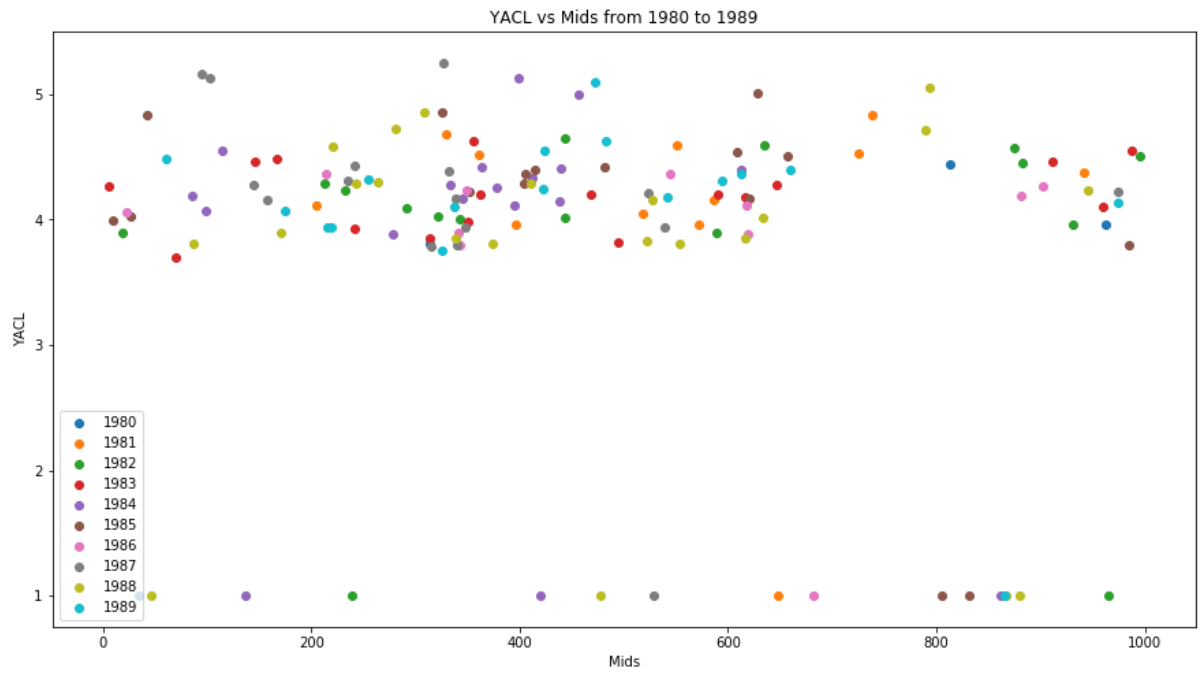
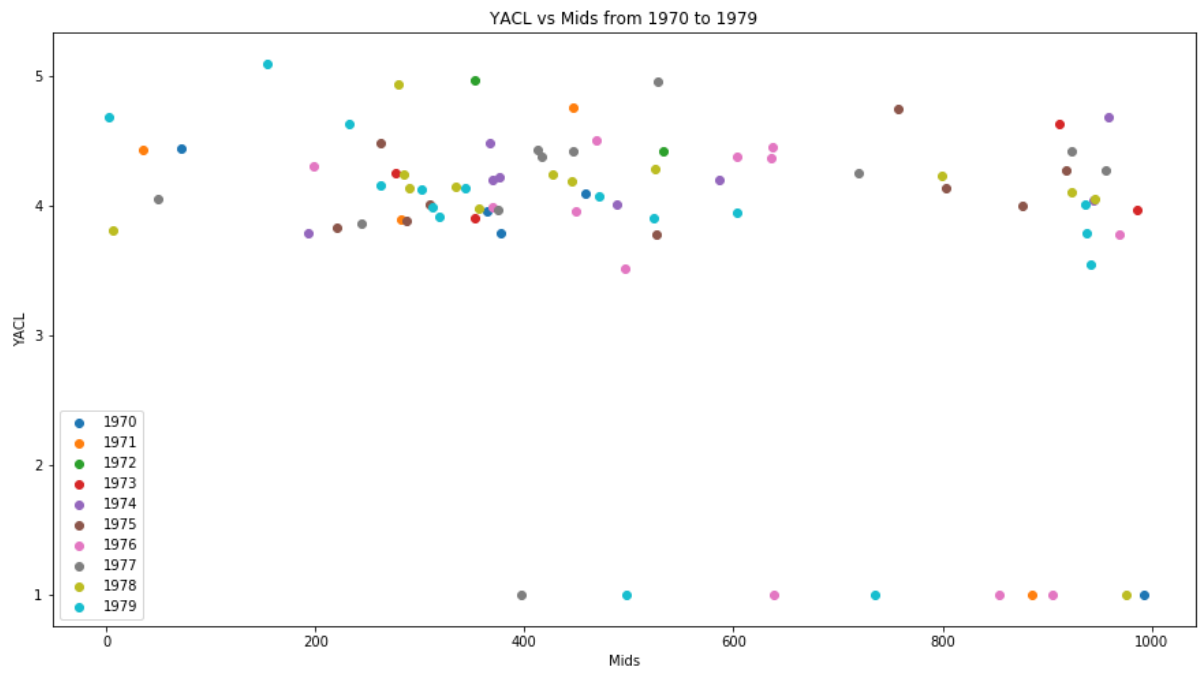
```
In [137]: def Build(yearLow, yearHigh):
    mov, yacIs = Run(yearLow,yearHigh);

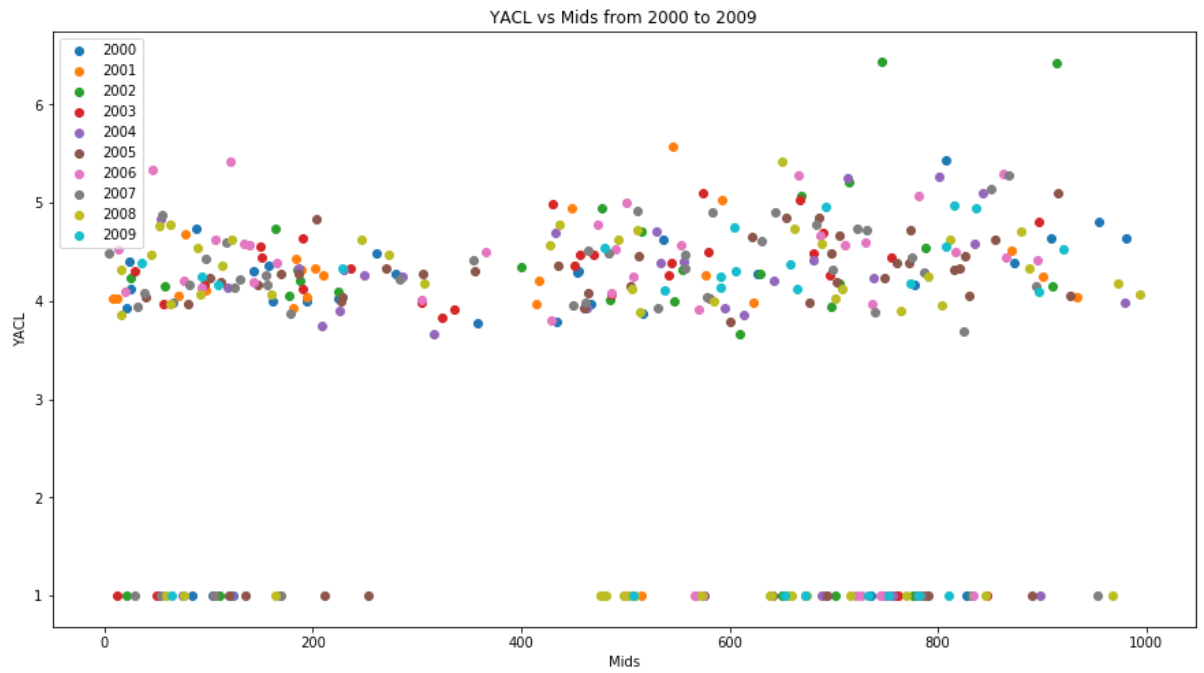
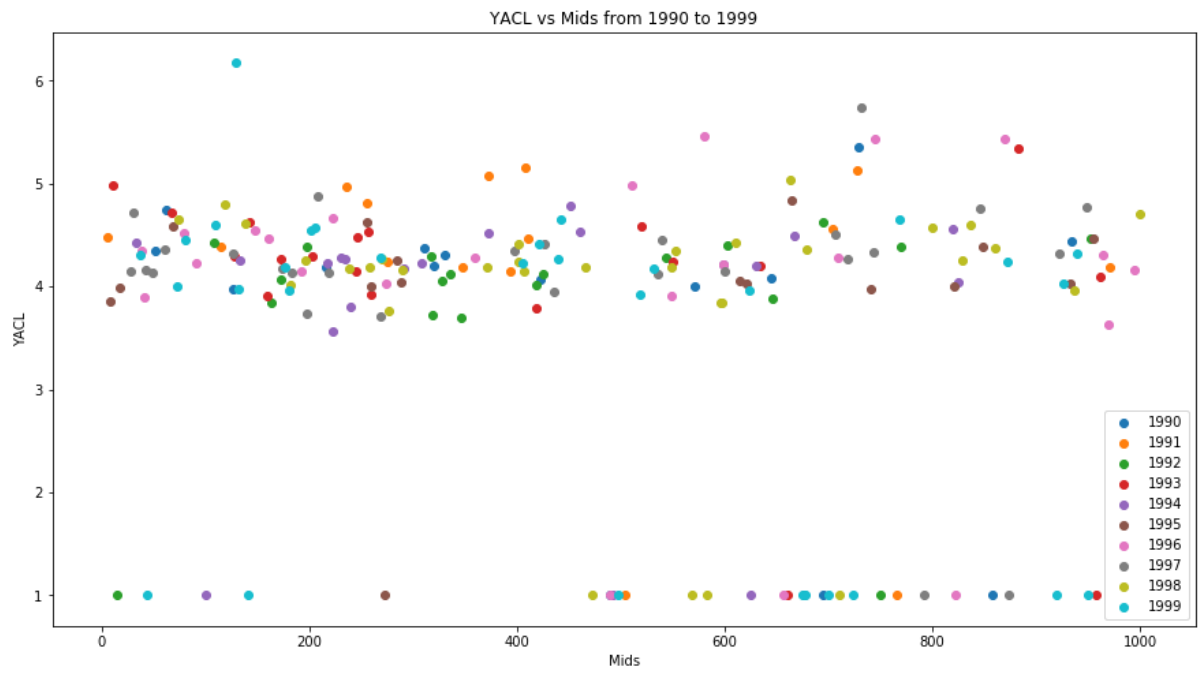
    plt.figure(figsize=(15,8))
    for i in range(yearHigh-yearLow):
        plt.scatter(mov[i],yacIs[i],marker='o',label=str(i+yearLow))
    plt.legend()
    plt.xlabel("Mids");
    plt.ylabel("YACL");
    plt.title("YACL vs Mids from "+str(yearLow)+" to "+str(yearHigh-1));
    plt.show()

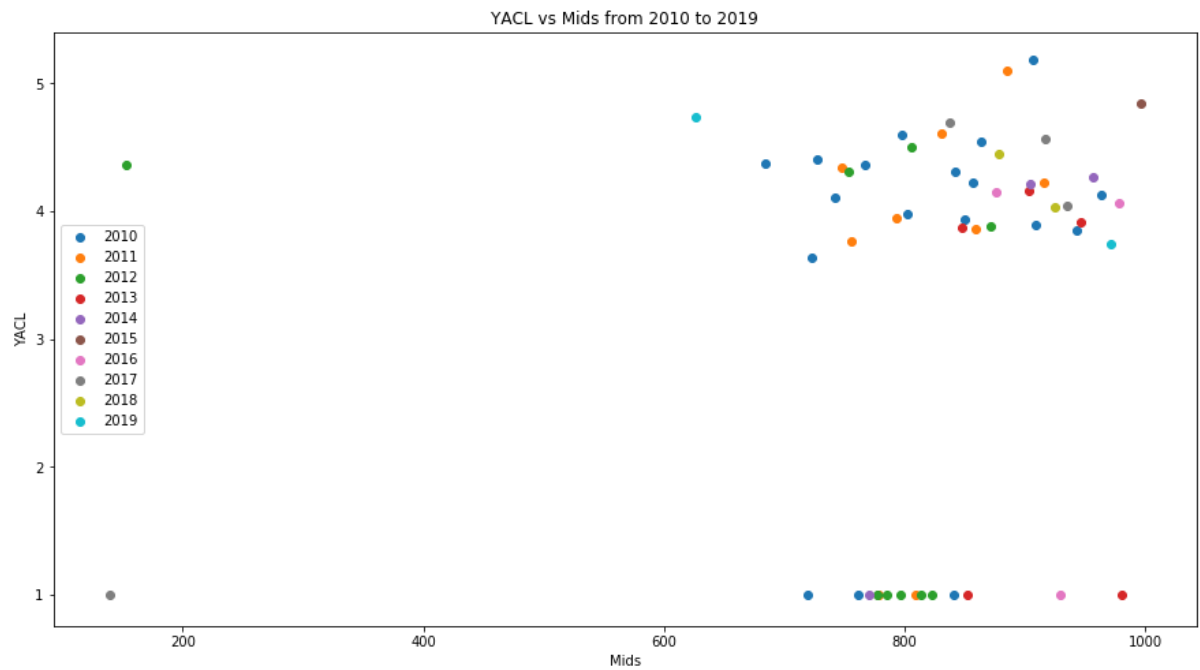
    yacIAvg = 0;
    for i in yacIs:
        if (len(i) > 0):
            yacIAvg += (sum(i)/len(i))
    yacIAvg = yacIAvg/len(yacIs)
    return yacIAvg;
```

```
In [140]: years = [1950,1960,1970,1980,1990,2000,2010]
          yacIs = [];
          for i in years:
              yacIs.append(Build(i,i+10));
```

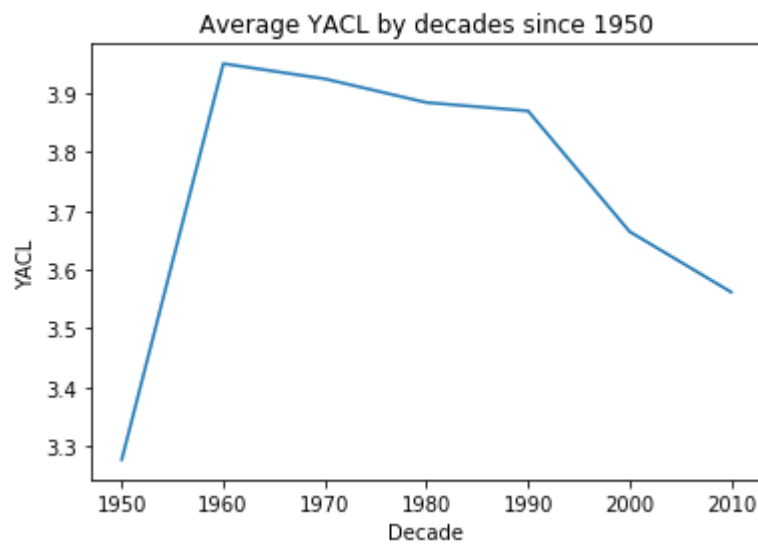








```
In [148]: plt.plot(years,yacIs)
plt.xlabel("Decade");
plt.ylabel("YACL");
plt.title("Average YACL by decades since 1950")
plt.show()
```



In []: