9.4 Distribution of Coefficients
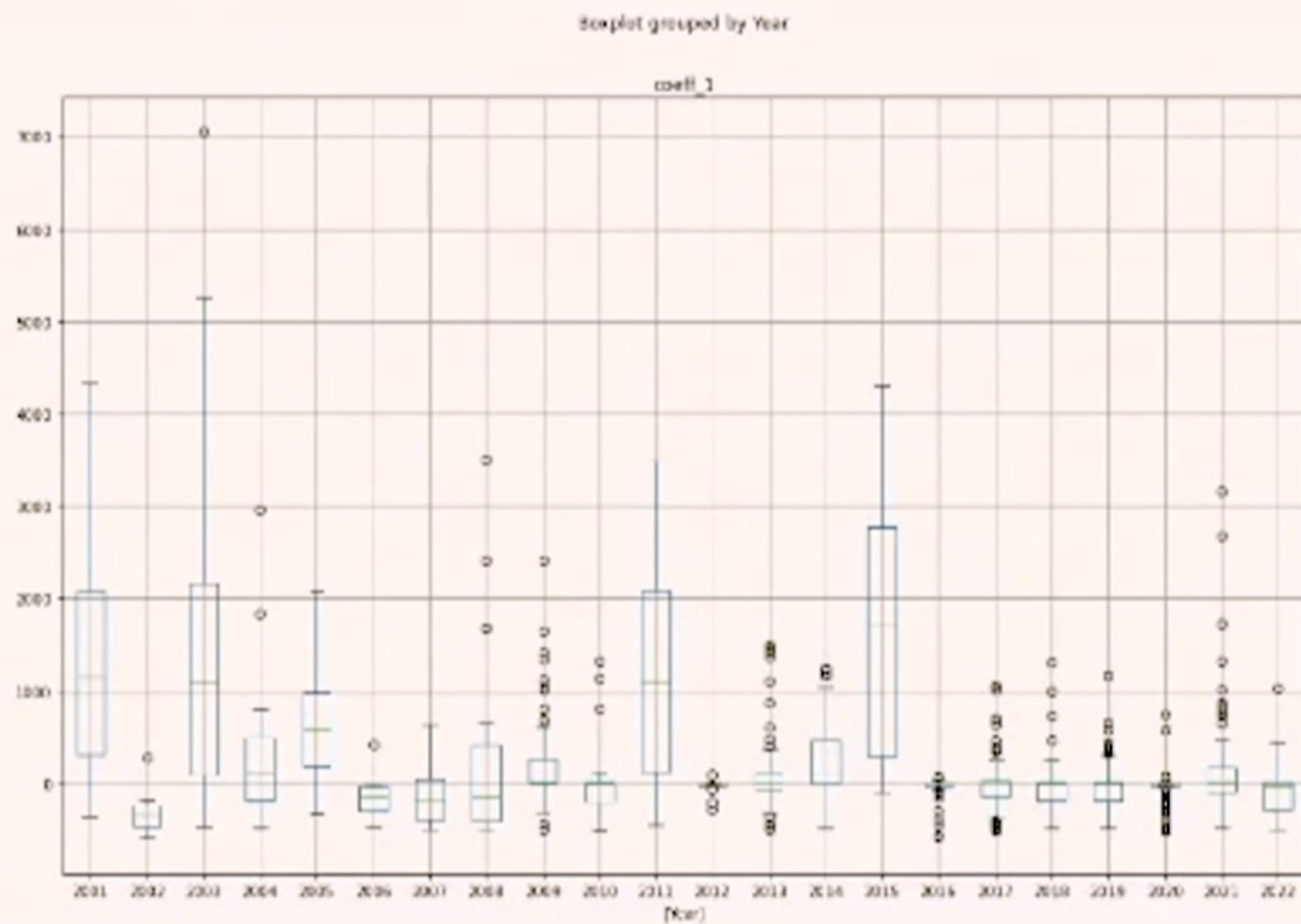
DSC 232R, Class 9: PCA for Weather Data

Studying the Distribution of the Coefficients

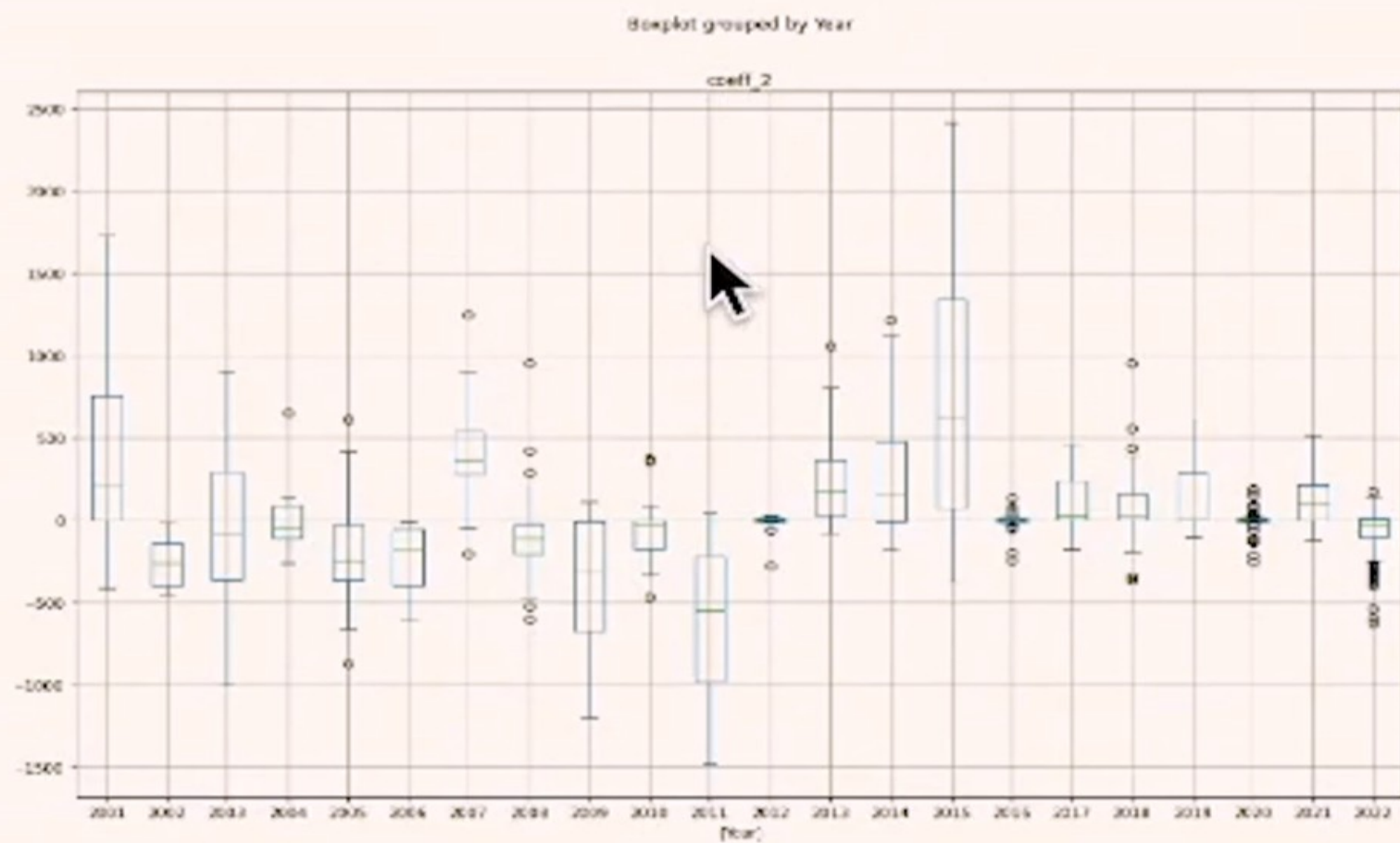Coeff_1-3 capture most of the variance in the snow-depth distribution

We can now look at how these coefficients vary from year to year

```
In [34]: pdf=df3.select(['Station','Year','coeff_1','coeff_2','coeff_3','res_1','res_2','res_3','res_mean','total_var']).toPandas()
```
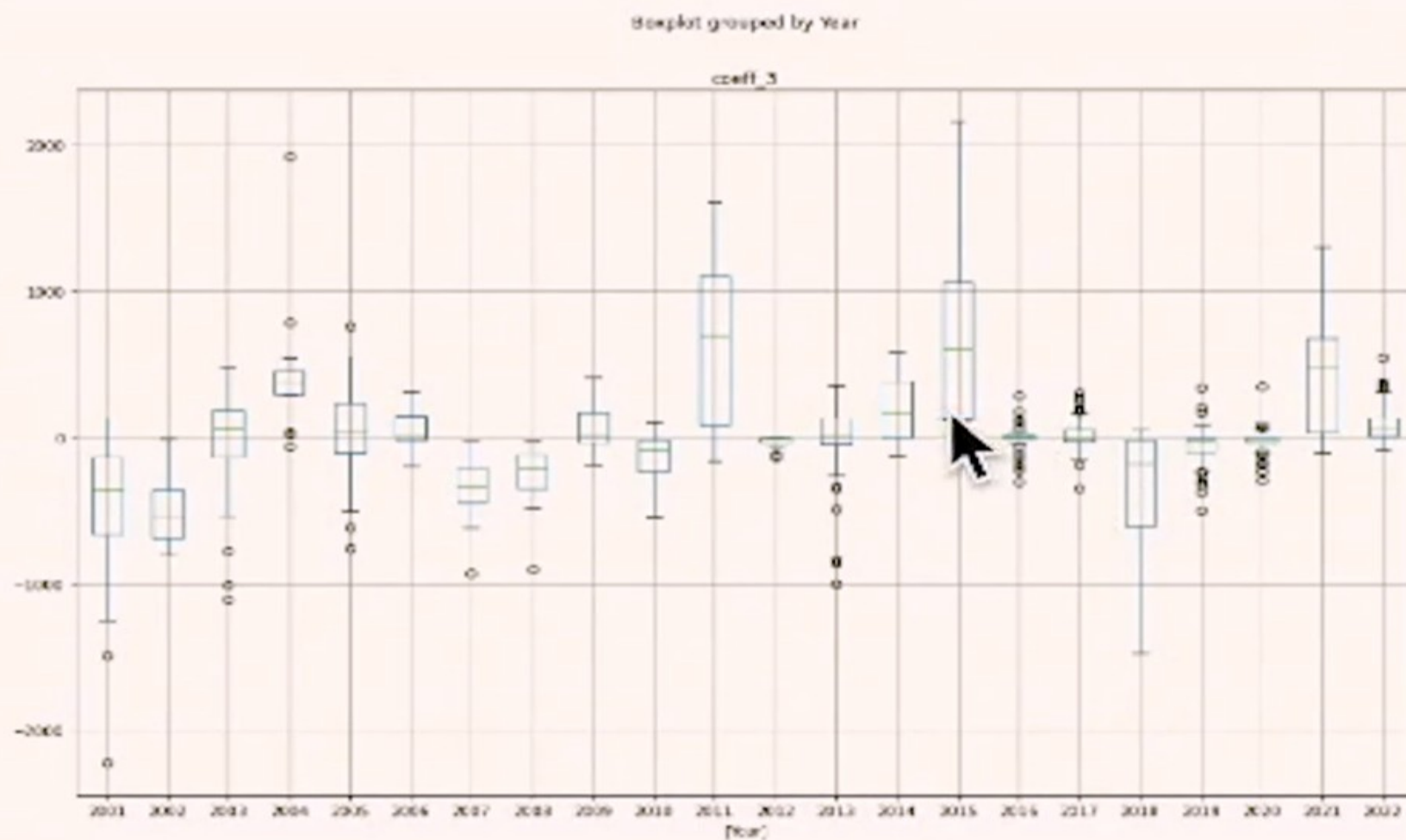
```
In [35]: #coeff_1 corresponds to the overall amount of snow.
         # 2001,2011, 2015 were heavy snow years
         pdf[['Year','coeff_1']][pdf['Year']>2000].boxplot(by='Year',figsize=[15,10]);
```


Boxplot grouped by Year

```
In [36]: #coeff_2 corresponds to whether the snow was early or late
         # 2011 was early 2015 was late
         pdf[['Year','coeff_2']][pdf['Year']>2000].boxplot(by='Year',figsize=[15,10]);
```
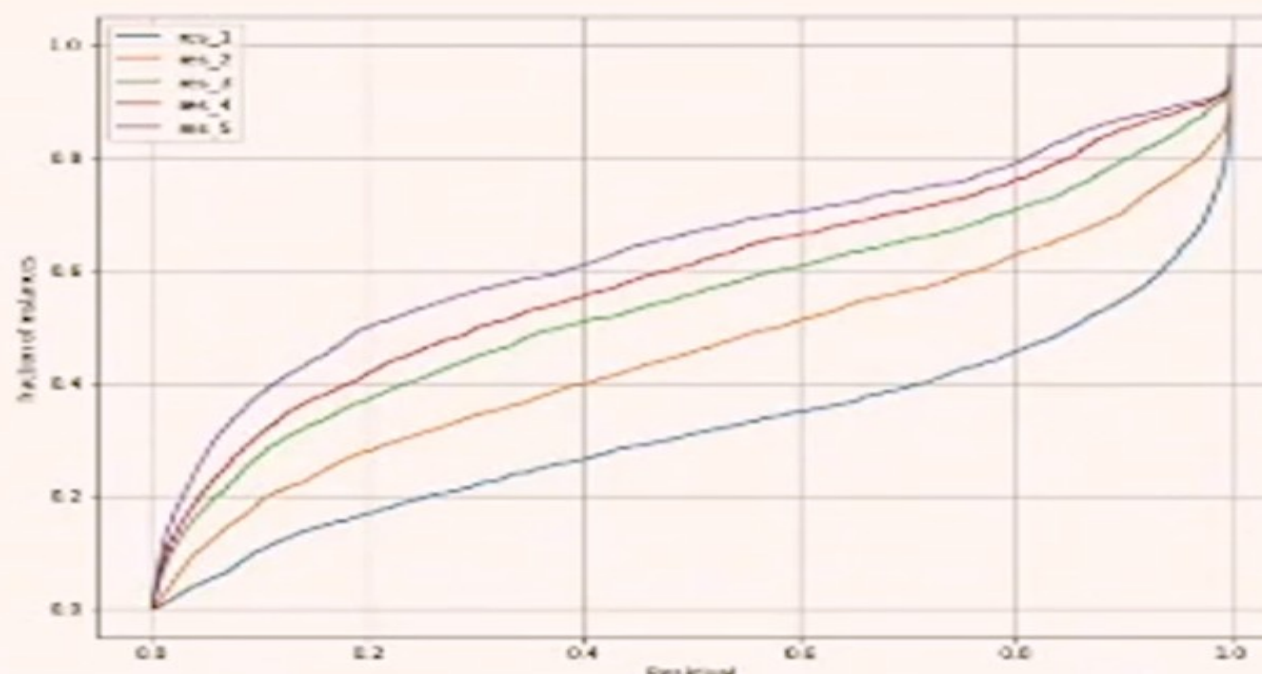

Boxplot grouped by Year

```
In [37]: #coeff_2 corresponds to whether the snow season was short or long
         # both 2011 and 2015 were long
         pdf[['Year','coeff_3']][pdf['Year']>2000].boxplot(by='Year',figsize=[15,10]);
```


Boxplot grouped by Year

```
In [40]: figure(figsize=(10,8))
         axis=gca()

         #plot_CDF('res_mean') # why does this not fit?
         plot_CDF('res_1')
         plot_CDF('res_2')
         plot_CDF('res_3')
         plot_CDF('res_4')
         plot_CDF('res_5')
         ylabel('fraction of instances')
         xlabel('Residual')
         grid()
         legend();
```
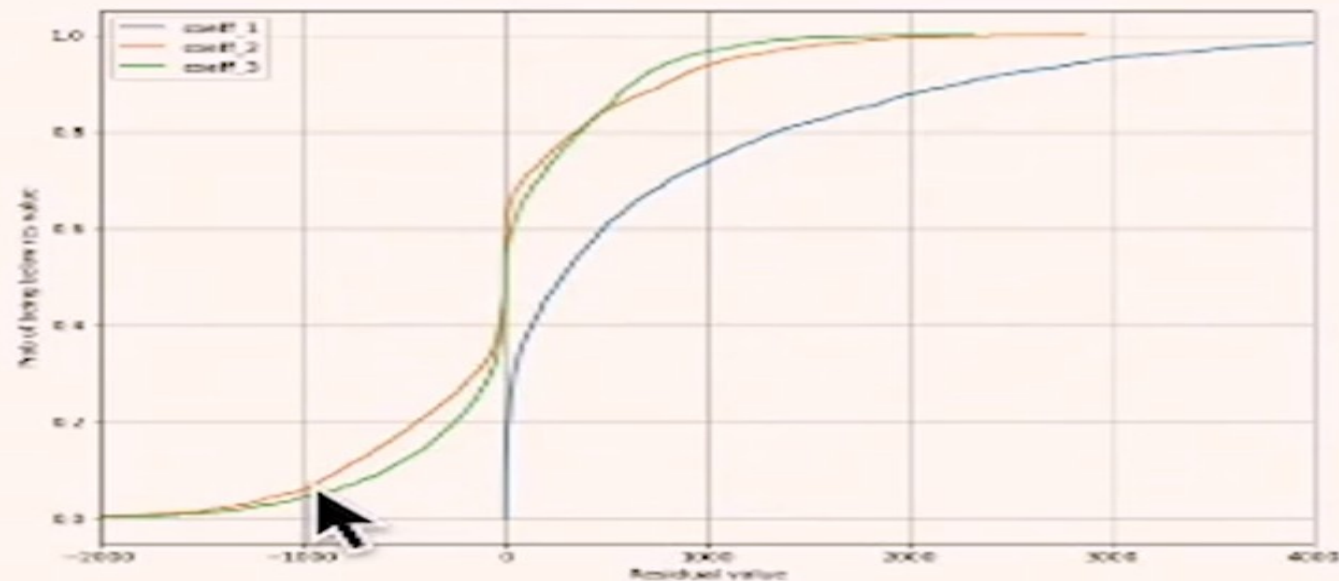
```
In [41]: figure(figsize=(10,8))
         axis=gca()

         plot_CDF('coeff_1',power=1)
         plot_CDF('coeff_2',power=1)
         plot_CDF('coeff_3',power=1)

         ylabel('Prob of being below res value')
         xlabel('Residual value')
         xlim([-2000,4000])
         grid()
         legend();
```



```
In [ ]:
```