Projects in Big Data Analytics

ISYE 4961/6961

Lecture 4

Project 1 – Using Wind Turbine Data

- Import files and extract some data columns into vectors
- Extract column segments
- Build some multi-column arrays
- Display and inspect the data (note: failure occurs at the start of data and time goes backward from there)
- Remove mean and scale, do some correlations and cross plots
- Look at groups of similar sensors, look at pairs of dissimilar sensors
- Look at different time blocks
- Look at same sensor across several turbines
- Report findings

Questions to Answer

- Which sensors are the most important?
- What are the distinguishing features of the data?
- Can we see any indications of failure in the raw data? After mean removal and scaling? In the histograms? In cross plots?
- What does normal operation look like and can we learned anything about regular operational performance?
- Ultimately, can we use the data to anticipate a failure?

Import from a file, extract a column

```
# read first 600 values from each input file data column into a list
>t1_1_600 <-read.table("all_outage_1_tab.txt",header=TRUE,nrows=600)

#look at first 10 values of each input data column
> t1_1_600[1:10,]

# extract a column from the list file into a vector
>col7 <- t1_1_600[[7]]

# extract multiple columns from the list file into a vector
>col789 <- c(col7,col8,col9)
>col789 <- array(col789, dim=c(3,500)
```

Scale and remove the mean

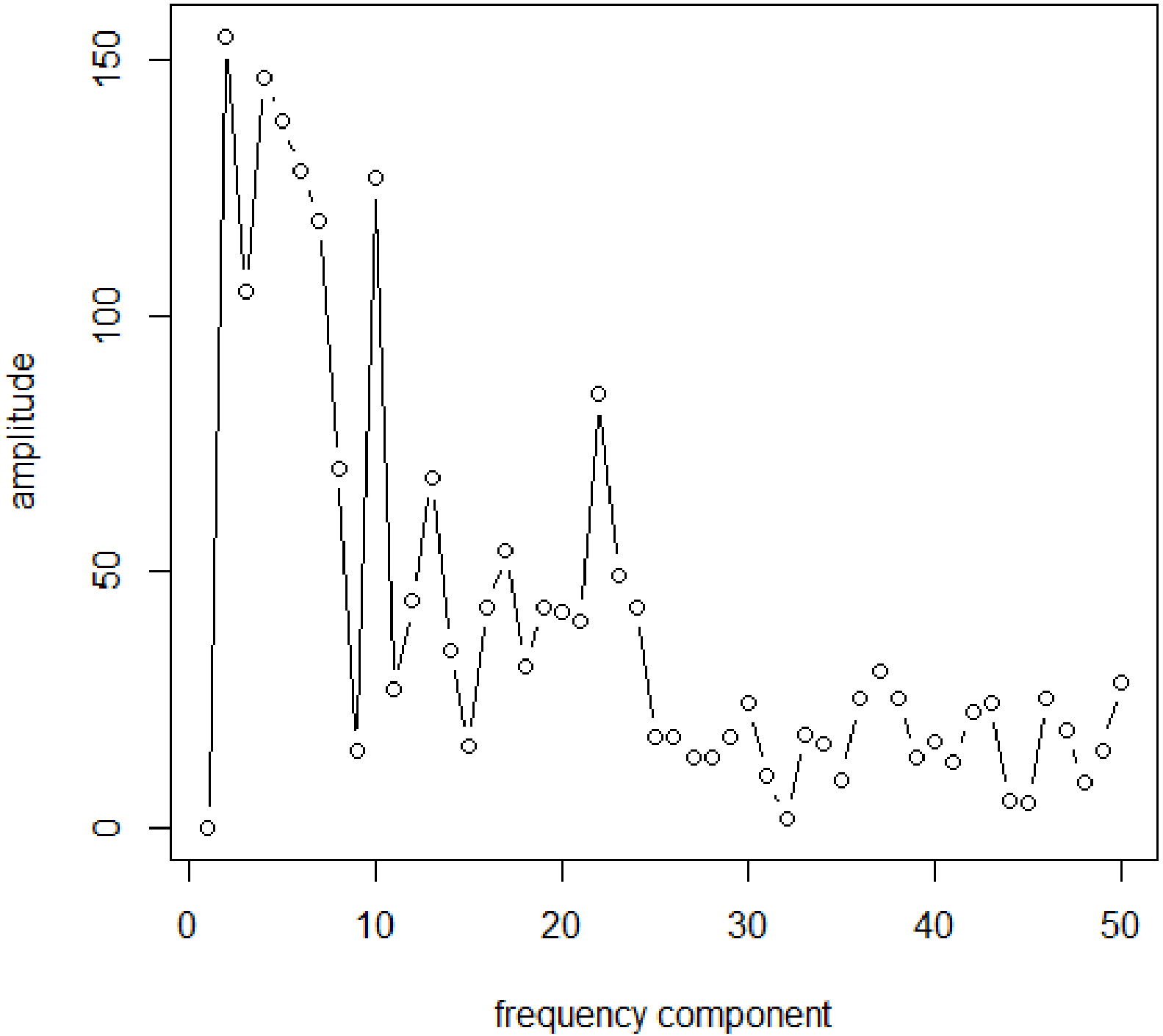
```
#convert to an array
col789_arr<-array(col789,dim=c(600,3))

#look at first 10 values of each input data column
> col789_arr[1:10,]

# remove the mean and scale
col789_arr_cs <- scale(col789_arr,center=TRUE,scale=TRUE)
```

Crossplot, calculate statistics

```
# crossplot pairs of data columns
>pairs(col789_arr_sc)



#calc covariances
cov(col789_arr_sc)
```

Fourier Transform

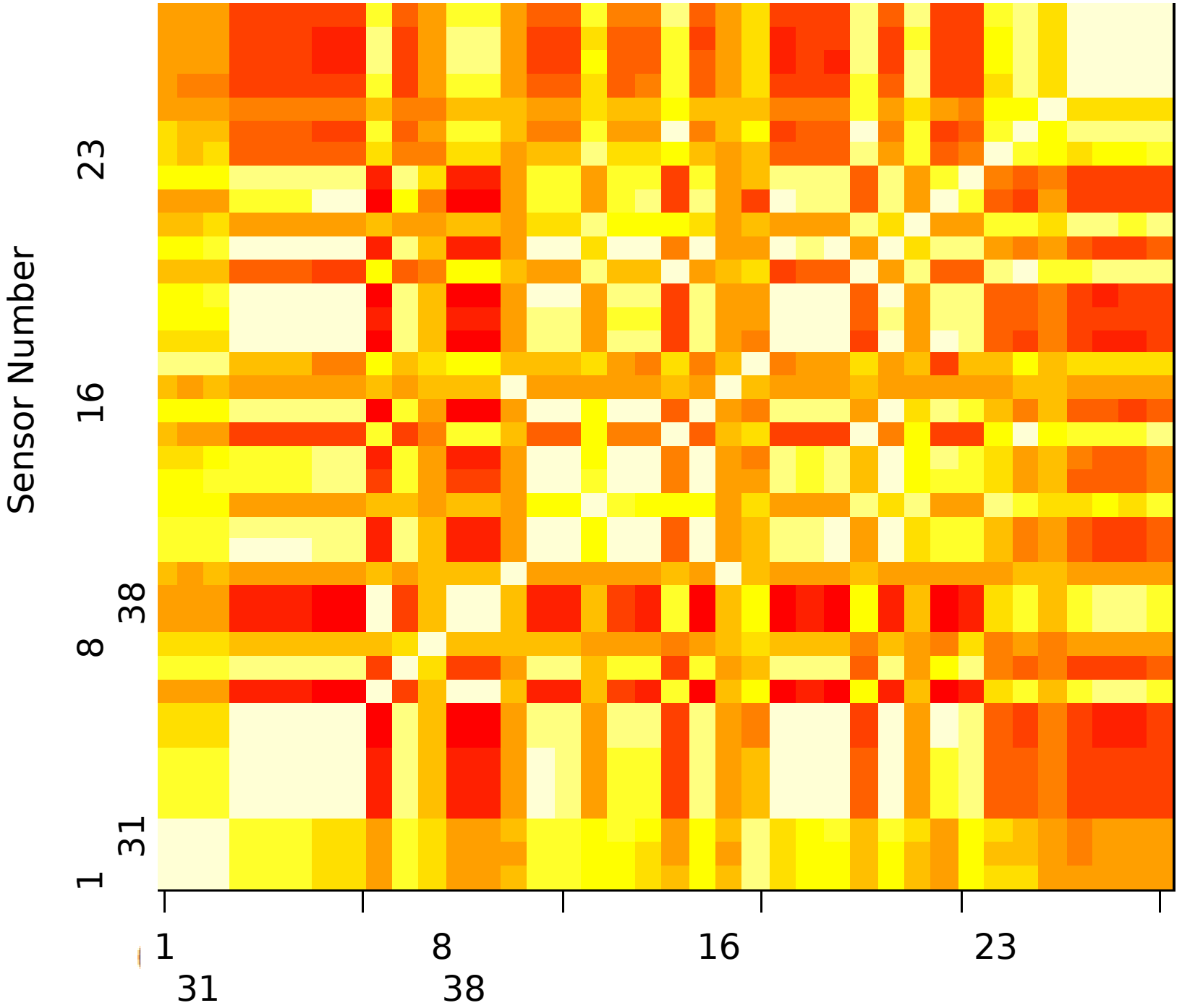
```
# transform the scaled and centered values of col 7
>col7_cs_f <- fft(col7_cs)
# calculate the amplitude (modulus or complex spectral magnitude)
>col7_cs_famp <- Mod(col7_cs_f)
# plot the results
>plot(col7_cs_famp[1:50],type=“b”, ,xlab="frequency component",ylab="amplitude")
```


amplitude
0
50
100
150
0
10
20
30
40
50
frequency component

Use scan to read the file

```
# scan file, skip the header row and read first 600 values
>t1_1_600 <-
scan("all_outage_1_tab.txt",skip=1,nlines=600)

#scan results in one long vector, convert to an array
t1_1_600_arr <- array(t1_1_600,dim=c(44,600))

#transpose the array
t1t_1_600_arr <- t(t1_1_600_arr)
```

```
# scale and center, skipping first 6 columns)
t1t_1_600_arr_sc <-
scale(t1t_1_600_arr[,7:44],center=TRUE,scale=TRUE)


#calculate covariances and save
T1_cov <- cov(t1t_1_600_arr_sc)


#display
Image(T1_cov)
```


Sensor Number
1
31
8
38
16
23
1
8
16
23
31
38