hw2

shichenh

9/13/2017

```
suppressMessages(library(tidyverse))
knitr::opts_chunk$set(echo = F)
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

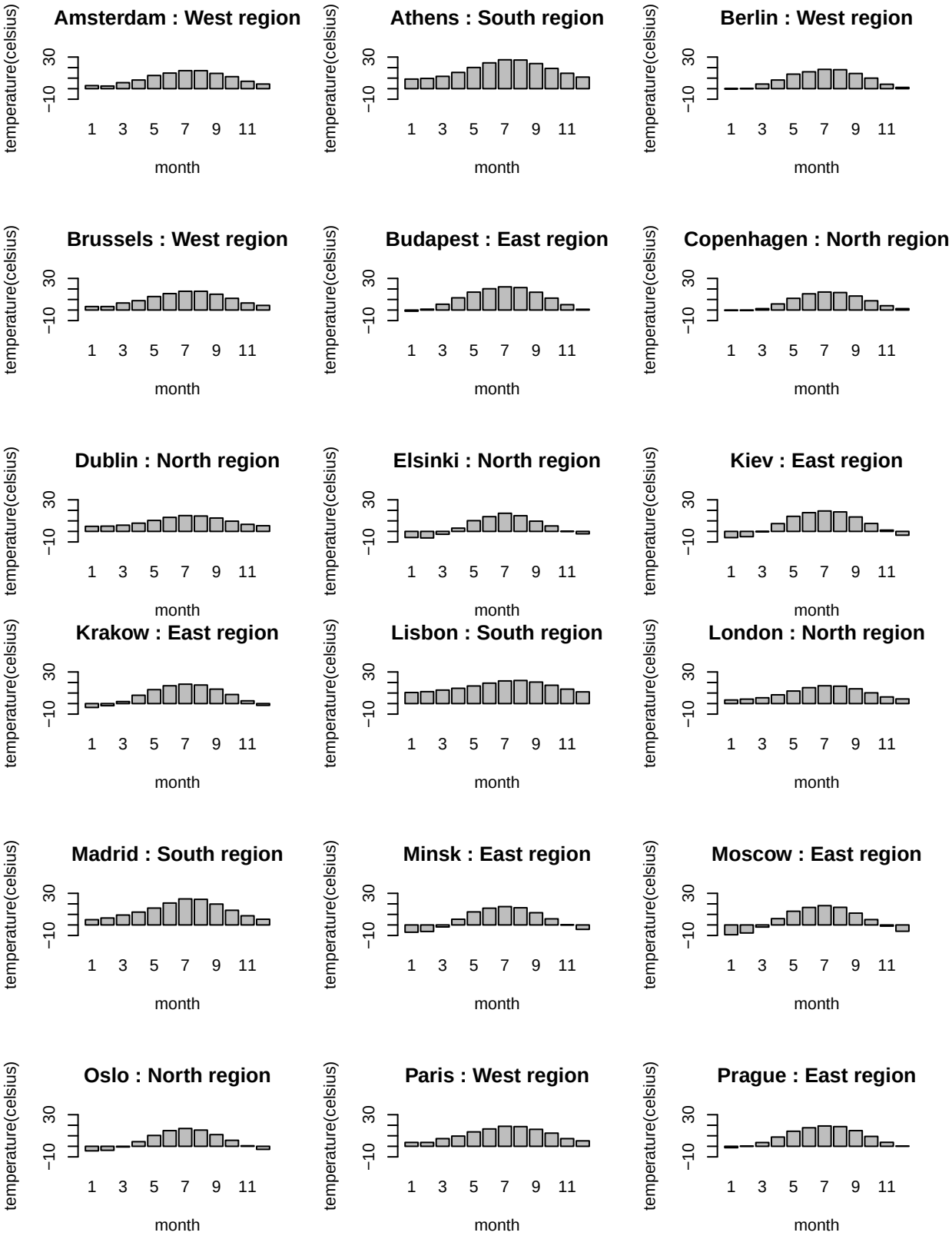
EDA1

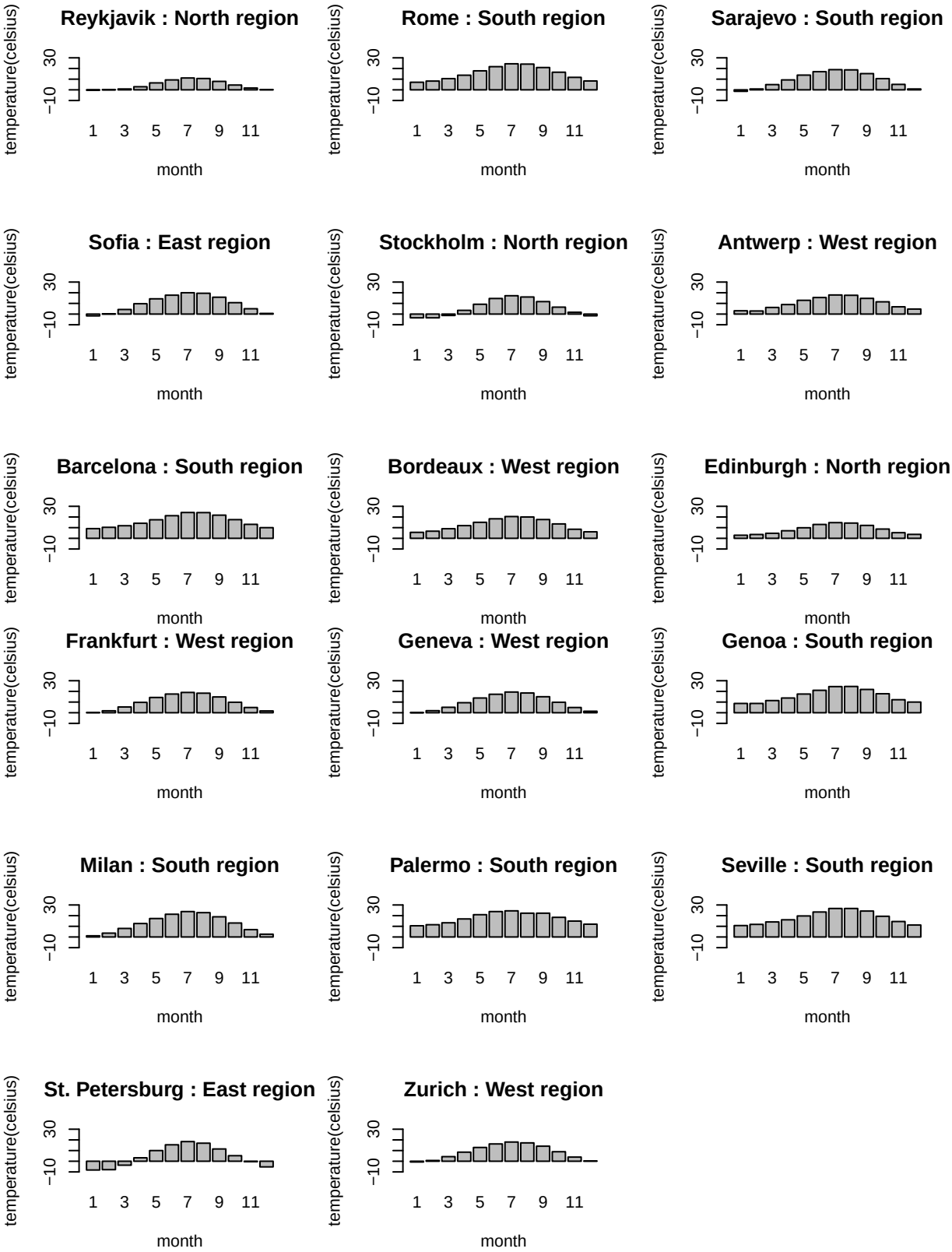
```
##    Amsterdam          Athens          Berlin          Brussels    
##  Min.   : 2.500   Min.   : 9.10   Min.   :-0.200   Min.   : 3.300  
##  1st Qu.: 5.375   1st Qu.:11.53   1st Qu.: 3.450   1st Qu.: 6.125  
##  Median : 9.800   Median :17.30   Median : 9.100   Median :10.000  
##  Mean   : 9.842   Mean   :17.81   Mean   : 9.033   Mean   :10.283  
##  3rd Qu.:14.575   3rd Qu.:23.98   3rd Qu.:14.800   3rd Qu.:15.150  
##  Max.   :17.100   Max.   :27.40   Max.   :18.300   Max.   :17.800  
##     Budapest        Copenhagen         Dublin          Elsinki      
##  Min.   :-1.100   Min.   :-0.400   Min.   : 4.800   Min.   :-6.200  
##  1st Qu.: 4.025   1st Qu.: 1.300   1st Qu.: 5.775   1st Qu.:-2.400  
##  Median :11.450   Median : 7.300   Median : 8.750   Median : 4.150  
##  Mean   :10.942   Mean   : 7.833   Mean   : 9.275   Mean   : 4.783  
##  3rd Qu.:17.800   3rd Qu.:13.825   3rd Qu.:12.850   3rd Qu.:11.150  
##  Max.   :22.000   Max.   :17.100   Max.   :15.000   Max.   :17.200  
##       Kiev            Krakow          Lisbon          London      
##  Min.   :-5.900   Min.   :-3.700   Min.   :10.50   Min.   : 3.400  
##  1st Qu.:-1.125   1st Qu.: 1.000   1st Qu.:12.43   1st Qu.: 5.225  
##  Median : 7.450   Median : 8.250   Median :15.60   Median : 9.250  
##  Mean   : 7.083   Mean   : 7.783   Mean   :15.93   Mean   : 9.725  
##  3rd Qu.:15.175   3rd Qu.:14.500   3rd Qu.:19.65   3rd Qu.:14.275  
##  Max.   :19.400   Max.   :18.400   Max.   :21.90   Max.   :16.900  
##      Madrid           Minsk           Moscow           Oslo        
##  Min.   : 5.000   Min.   :-6.900   Min.   :-9.300   Min.   :-4.300  
##  1st Qu.: 8.175   1st Qu.:-2.475   1st Qu.:-3.000   1st Qu.:-1.175  
##  Median :13.050   Median : 5.600   Median : 5.550   Median : 5.050  
##  Mean   :13.900   Mean   : 5.475   Mean   : 5.075   Mean   : 5.633  
##  3rd Qu.:20.050   3rd Qu.:13.275   3rd Qu.:13.900   3rd Qu.:12.050  
##  Max.   :24.700   Max.   :17.400   Max.   :18.300   Max.   :16.900  
##      Paris           Prague         Reykjavik          Rome        
##  Min.   : 3.700   Min.   :-1.300   Min.   :-0.300   Min.   : 7.10  
##  1st Qu.: 6.775   1st Qu.: 2.775   1st Qu.: 0.650   1st Qu.: 9.95  
##  Median :11.100   Median : 9.100   Median : 3.700   Median :15.10  
##  Mean   :11.117   Mean   : 9.133   Mean   : 4.608   Mean   :15.41  
##  3rd Qu.:16.200   3rd Qu.:15.575   3rd Qu.: 8.250   3rd Qu.:21.10  
##  Max.   :19.000   Max.   :19.300   Max.   :11.100   Max.   :24.40  
##     Sarajevo          Sofia         Stockholm         Antwerp     
##  Min.   :-1.400   Min.   :-1.700   Min.   :-3.500   Min.   : 2.900  
##  1st Qu.: 3.875   1st Qu.: 3.375   1st Qu.:-1.375   1st Qu.: 5.825  
##  Median : 9.900   Median :10.200   Median : 5.000   Median :10.200
```

```
##  Mean   : 9.467   Mean   : 9.675   Mean   : 5.875   Mean   :10.225
##  3rd Qu.:15.650   3rd Qu.:16.275   3rd Qu.:12.425   3rd Qu.:14.900
##  Max.   :18.900   Max.   :20.000   Max.   :17.200   Max.   :17.900
##    Barcelona        Bordeaux       Edinburgh        Frankfurt
##  Min.   : 9.10   Min.   : 5.60   Min.   : 2.900   Min.   : 0.200
##  1st Qu.:11.43   1st Qu.: 8.05   1st Qu.: 4.450   1st Qu.: 4.125
##  Median :15.75   Median :12.70   Median : 7.900   Median : 9.750
##  Mean   :16.21   Mean   :12.72   Mean   : 8.333   Mean   : 9.783
##  3rd Qu.:21.32   3rd Qu.:17.77   3rd Qu.:12.325   3rd Qu.:15.475
##  Max.   :24.20   Max.   :20.40   Max.   :14.700   Max.   :19.000
##      Geneva           Genoa           Milan           Palermo
##  Min.   : 0.100   Min.   : 8.70   Min.   : 1.100   Min.   :10.50
##  1st Qu.: 4.150   1st Qu.:11.05   1st Qu.: 6.075   1st Qu.:12.97
##  Median : 9.600   Median :15.65   Median :12.850   Median :17.65
##  Mean   : 9.717   Mean   :16.00   Mean   :12.667   Mean   :17.61
##  3rd Qu.:15.575   3rd Qu.:21.20   3rd Qu.:19.500   3rd Qu.:22.30
##  Max.   :19.400   Max.   :24.60   Max.   :23.800   Max.   :24.50
##     Seville      St. Petersburg       Zurich
##  Min.   :10.70   Min.   :-8.200   Min.   :-0.700
##  1st Qu.:13.53   1st Qu.:-4.100   1st Qu.: 3.100
##  Median :17.75   Median : 4.200   Median : 8.700
##  Mean   :18.22   Mean   : 4.592   Mean   : 8.692
##  3rd Qu.:23.62   3rd Qu.:12.475   3rd Qu.:14.625
##  Max.   :26.70   Max.   :18.400   Max.   :18.000
```

Summary: The cities do not really have a super high max temperature, at the same tiem many have mean temperature lower than 10(pretty cold XD)

```
##
##  East North South  West
##     8     8    10     9
```


Amsterdam : West region
Athens : South region
Berlin : West region
Brussels : West region
Budapest : East region
Copenhagen : North region
Dublin : North region
Elsinki : North region
Kiev : East region
Krakow : East region
Lisbon : South region
London : North region
Madrid : South region
Minsk : East region
Moscow : East region
Oslo : North region
Paris : West region
Prague : East region
temperature(celsius)
−10
30
1 3 5 7 9 11
month



Seems like East region and north region tend to have temperatures during the winter. I wonder why not the West case? Maybe because the west region is closer to the ocean so it makes the cities warmer?

PCA

raw

1a displaying first 4 loadings

```
##             [,1]        [,2]          [,3]         [,4]
##  [1,] -0.2671050 -0.39091041  0.1907187341 -0.059731884
##  [2,] -0.2803688 -0.33534791 -0.0097552190 -0.427798846
##  [3,] -0.2996355 -0.21137095 -0.3399569587 -0.397667051
##  [4,] -0.3087780  0.07324821 -0.5579573828 -0.127078736
##  [5,] -0.2757927  0.33680390 -0.4392770157  0.392591602
##  [6,] -0.2642082  0.40118372  0.1394431457 -0.000489339
##  [7,] -0.2676478  0.37421361  0.4325313064 -0.222824851
##  [8,] -0.2882824  0.29568869  0.2462557102 -0.226852869
##  [9,] -0.3124996  0.11221817  0.0636774480 -0.026537477
## [10,] -0.3144017 -0.06235990 -0.0001874864  0.366581807
## [11,] -0.3019515 -0.21291689  0.1244515912  0.356372148
## [12,] -0.2768287 -0.34787886  0.2386777766  0.349002937
```

1b obtaining principal components

```
##             [,1]         [,2]        [,3]        [,4]         [,5]
##  [1,] -0.22195025 -1.341234829 -0.10209889  0.27657677  0.220284157
##  [2,] -7.43360390  0.909925426  0.54908835  0.28025851 -0.118618176
##  [3,]  0.28153099  0.016092403 -0.28422057  0.05437108  0.138661074
##  [4,] -0.61729994 -1.151341565 -0.14870076 -0.01669466  0.115484812
##  [5,] -1.63136395  1.675051425 -0.48801530 -0.10996512 -0.144251778
##  [6,]  1.43025066 -0.481240562  0.43068897  0.17283180 -0.001329907
##  [7,]  0.49413580 -2.614731574 -0.17458563 -0.02925371 -0.195104195
##  [8,]  3.94757646  0.451883416  0.58015037  0.23907168 -0.057464374
##  [9,]  1.67458427  1.963469194 -0.16691889  0.11032784 -0.045733930
## [10,]  1.23099109  0.855756199 -0.26794138 -0.03573418  0.021782597
## [11,] -5.47621202 -1.520180219 -0.26440940  0.13422375 -0.056722627
## [12,] -0.05637309 -1.539174219 -0.08281278 -0.05087152 -0.162725820
## [13,] -3.97473636  0.682329696  0.45164881 -0.64836153  0.075052049
## [14,]  3.16672621  1.360708200 -0.07068160  0.17931195 -0.065426231
## [15,]  3.38650106  2.134053560 -0.29467958  0.00526448 -0.074113520
## [16,]  3.23331905  0.303237840  0.28881834 -0.18641912 -0.067051552
## [17,] -1.38850720 -0.877868695 -0.10790241  0.07732927  0.180820308
## [18,]  0.10660691  0.682697725 -0.23723947 -0.09816888 -0.117176932
## [19,]  4.60066569 -2.892196405 -0.05662577 -0.19107214 -0.100488470
## [20,] -5.26370105  0.287243017  0.18510843  0.01231239 -0.075942846
## [21,] -0.15985914  0.312466849 -0.35657228 -0.07199691  0.200165679
## [22,] -0.40862719  0.777598162 -0.23556939 -0.04675281  0.190733740
## [23,]  3.07934588  0.005454959  0.85347084 -0.05658895  0.139165944
##                [,6]         [,7]          [,8]         [,9]        [,10]
##  [1,]  0.0606515206  0.013837928 -3.551450e-02 -0.001258429 -0.027524896
##  [2,] -0.0214851576 -0.009477015 -3.470022e-02 -0.021689471 -0.026321973
##  [3,]  0.1784390246  0.156272684 -1.939858e-02 -0.039940006  0.024012803
##  [4,]  0.0952662184  0.044341741  1.536379e-02  0.031685340 -0.069718521
##  [5,]  0.0603055393  0.003433762  2.767219e-02  0.005731185 -0.004984876
```

```
##  [6,] -0.1846972943  0.099340907  9.178098e-03  0.004371045  0.025801813
##  [7,] -0.0094018948 -0.061657515  3.185013e-02  0.048187791 -0.004986983
##  [8,]  0.2938787670 -0.135338158  4.251743e-03 -0.009838692  0.044058991
##  [9,] -0.0469943751  0.098135394 -4.670525e-02  0.057784139  0.005086755
## [10,] -0.1457866006 -0.026245745  6.035049e-02 -0.018337188  0.029948174
## [11,] -0.0753226346 -0.055443872 -2.235481e-03  0.021915104  0.002733916
## [12,] -0.0283374480  0.008579577  1.300830e-02 -0.008782549  0.036659603
## [13,]  0.1149861566  0.008636404 -4.315413e-02  0.032253936  0.009046625
## [14,] -0.0340541573  0.048549857 -4.859301e-02  0.023586240 -0.018244925
## [15,]  0.0520824066 -0.119037818  2.123921e-05  0.004421133 -0.045965015
## [16,] -0.0282648878  0.044074185  9.294155e-02 -0.033148590 -0.048742710
## [17,]  0.1162195804  0.027251024  9.976817e-02  0.024382612  0.022767597
## [18,] -0.0039540687  0.106156819 -2.543791e-03 -0.028622521  0.032408636
## [19,] -0.0043869032 -0.008913707 -9.699846e-02 -0.032814430 -0.005907064
## [20,] -0.0009733389 -0.012884873  9.642558e-03 -0.041805177 -0.002396962
## [21,] -0.1183861065 -0.105677816 -1.920387e-02 -0.066782214 -0.015878743
## [22,] -0.1095359855 -0.118901670 -2.916616e-02  0.032785611  0.047796271
## [23,] -0.1602483605 -0.005032092  1.416518e-02  0.015915132 -0.009648517
##               [,11]         [,12]
##  [1,] -0.0311545855  0.0062040715
##  [2,]  0.0183196936 -0.0014967996
##  [3,] -0.0206419478  0.0006844792
##  [4,]  0.0373296324  0.0035175614
##  [5,] -0.0146505861  0.0223015596
##  [6,]  0.0223244352  0.0232279983
##  [7,]  0.0003430448  0.0004004023
##  [8,] -0.0008939795  0.0006955612
##  [9,] -0.0218730319 -0.0054785873
## [10,] -0.0261095773  0.0179264748
## [11,] -0.0323656812 -0.0092223468
## [12,]  0.0285970649 -0.0005603338
## [13,] -0.0007772308  0.0072068881
## [14,]  0.0129816117 -0.0097753345
## [15,]  0.0140655248  0.0068262935
## [16,] -0.0319315179 -0.0223920841
## [17,]  0.0222121748 -0.0047208796
## [18,]  0.0369542347 -0.0181226369
## [19,] -0.0124083260  0.0025004092
## [20,] -0.0231282989 -0.0024175070
## [21,]  0.0308859351  0.0013023776
## [22,] -0.0006498259 -0.0190821456
## [23,] -0.0074287632  0.0004745785
```

1c display eigen value

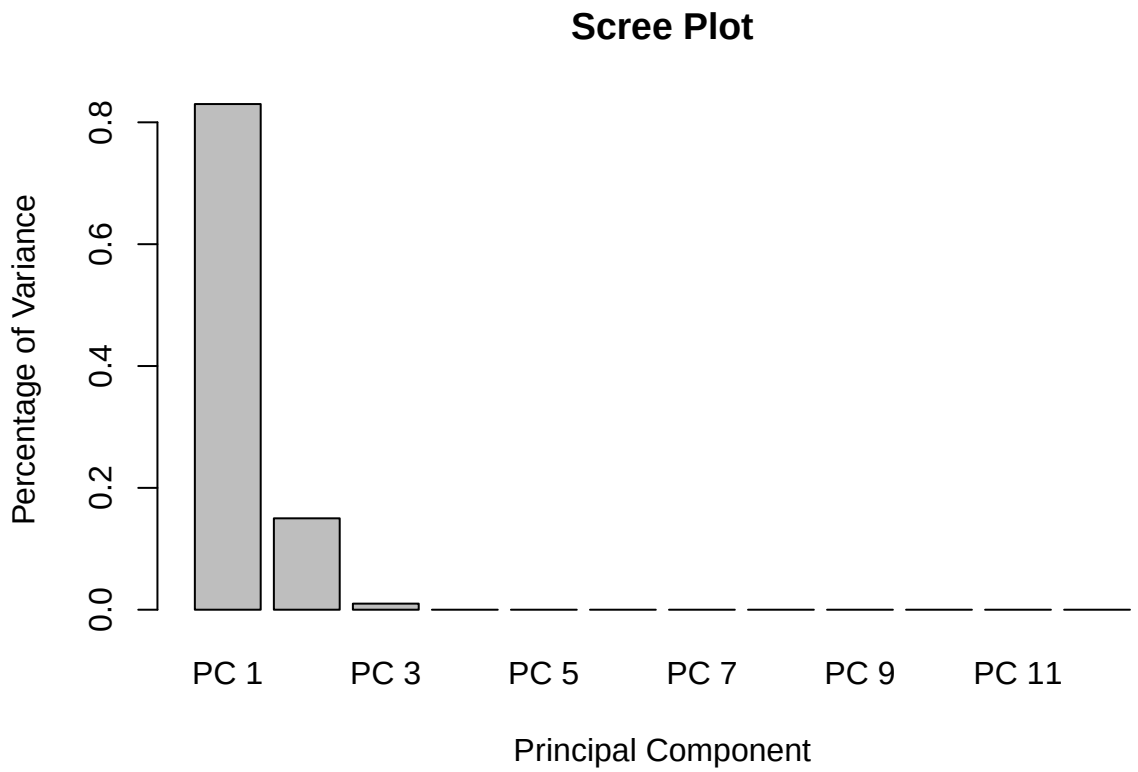
```
## [1] "eigen values"
##  [1] 9.95 1.85 0.13 0.04 0.02 0.01 0.01 0.00 0.00 0.00 0.00 0.00
## [1] "sum of eigen values"
## [1] 12
```

2

2a

```
##            var percent percent.cumulative
## 1  9.9477504204    0.83          0.8289792
## 2  1.8476485015    0.15          0.9829499
## 3  0.1262558038    0.01          0.9934712
## 4  0.0382934463    0.00          0.9966623
## 5  0.0167094089    0.00          0.9980548
## 6  0.0128330357    0.00          0.9991242
## 7  0.0058302931    0.00          0.9996101
## 8  0.0020318929    0.00          0.9997794
## 9  0.0010234516    0.00          0.9998647
## 10 0.0009527707    0.00          0.9999441
## 11 0.0005367834    0.00          0.9999888
## 12 0.0001341917    0.00          1.0000000
```

2b



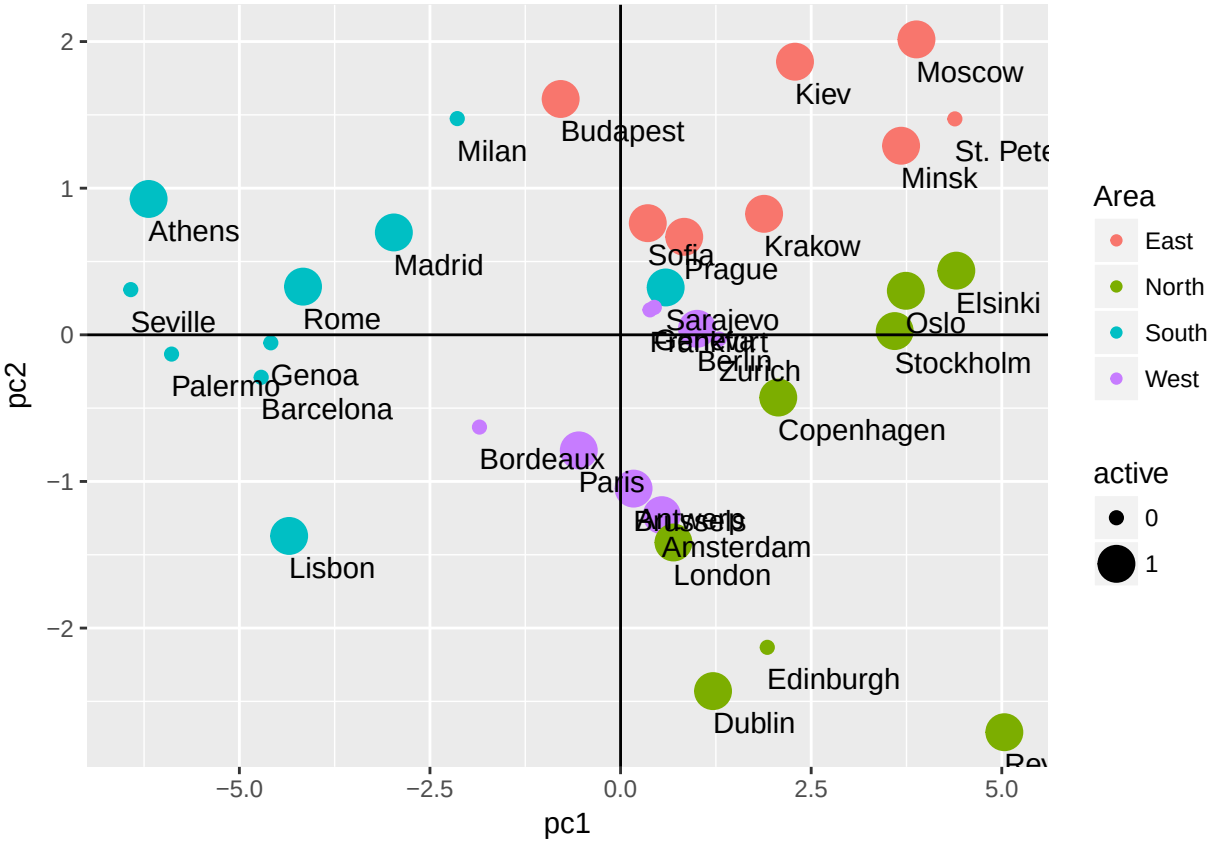
b.From the chart, it seems only two(probably3) PC contributes to significant amount of variance. In other words, many data are redundant.

c.I would choose the firs 3 dimensions, because it already contains 99% of the data variation.

3

3a making a scatter plot

```
## Warning: Using size for a discrete variable is not advised.
```



Most of the South Region cities are in the negative area of PC1, spread in pc2. The North region cities are mostly in the positive axes of pc1. The West region cities are mostly in the negative axes of city 2. Most of the East region cities are in the positive of both axes.

3b square cosine

```
##         [,1]   [,2]   [,3]   [,4]
##  [1,] 0.0247 0.9037 0.0052 0.0384
##  [2,] 0.9783 0.0147 0.0053 0.0014
##  [3,] 0.3279 0.0011 0.3342 0.0122
##  [4,] 0.2164 0.7528 0.0126 0.0002
##  [5,] 0.4634 0.4885 0.0415 0.0021
##  [6,] 0.8059 0.0912 0.0731 0.0118
##  [7,] 0.0341 0.9552 0.0043 0.0001
##  [8,] 0.9565 0.0125 0.0207 0.0035
##  [9,] 0.4173 0.5737 0.0041 0.0018
## [10,] 0.6451 0.3118 0.0306 0.0005
## [11,] 0.9255 0.0713 0.0022 0.0006
## [12,] 0.0013 0.9824 0.0028 0.0011
## [13,] 0.9342 0.0275 0.0121 0.0249
```

```
## [14,] 0.8407 0.1552 0.0004 0.0027
## [15,] 0.7108 0.2823 0.0054 0.0000
## [16,] 0.9784 0.0086 0.0078 0.0033
## [17,] 0.6948 0.2777 0.0042 0.0022
## [18,] 0.0199 0.8149 0.0984 0.0168
## [19,] 0.7153 0.2827 0.0001 0.0012
## [20,] 0.9955 0.0030 0.0012 0.0000
## [21,] 0.0782 0.2988 0.3891 0.0159
## [22,] 0.1863 0.6745 0.0619 0.0024
## [23,] 0.9242 0.0000 0.0710 0.0003
```

3 c

```
##             PC 1  PC 2  PC 3  PC 4  PC 5  PC 6  PC 7  PC 8  PC 9 PC 10
## Amsterdam   0.02  4.43  0.38  9.08 13.20  1.30  0.15  2.82  0.01  3.61
## Athens     25.25  2.04 10.85  9.32  3.83  0.16  0.07  2.69  2.09  3.31
## Berlin      0.04  0.00  2.91  0.35  5.23 11.28 19.04  0.84  7.08  2.75
## Brussels    0.17  3.26  0.80  0.03  3.63  3.21  1.53  0.53  4.46 23.19
## Budapest    1.22  6.90  8.57  1.44  5.66  1.29  0.01  1.71  0.15  0.12
## Copenhagen  0.93  0.57  6.68  3.55  0.00 12.08  7.69  0.19  0.08  3.18
## Dublin      0.11 16.82  1.10  0.10 10.35  0.03  2.96  2.27 10.31  0.12
## Elsinki     7.12  0.50 12.12  6.78  0.90 30.59 14.28  0.04  0.43  9.26
## Kiev        1.28  9.48  1.00  1.44  0.57  0.78  7.51  4.88 14.83  0.12
## Krakow      0.69  1.80  2.58  0.15  0.13  7.53  0.54  8.15  1.49  4.28
## Lisbon     13.70  5.69  2.52  2.14  0.88  2.01  2.40  0.01  2.13  0.04
## London      0.00  5.83  0.25  0.31  7.20  0.28  0.06  0.38  0.34  6.41
## Madrid      7.22  1.15  7.34 49.90  1.53  4.68  0.06  4.17  4.62  0.39
## Minsk       4.58  4.55  0.18  3.82  1.16  0.41  1.84  5.28  2.47  1.59
## Moscow      5.24 11.20  3.13  0.00  1.49  0.96 11.05  0.00  0.09 10.08
## Oslo        4.78  0.23  3.00  4.13  1.22  0.28  1.51 19.32  4.88 11.33
## Paris       0.88  1.90  0.42  0.71  8.89  4.78  0.58 22.27  2.64  2.47
## Prague      0.01  1.15  2.03  1.14  3.74  0.01  8.79  0.01  3.64  5.01
## Reykjavik   9.67 20.58  0.12  4.33  2.75  0.01  0.06 21.05  4.78  0.17
## Rome       12.66  0.20  1.23  0.02  1.57  0.00  0.13  0.21  7.76  0.03
## Sarajevo    0.01  0.24  4.58  0.62 10.90  4.96  8.71  0.82 19.81  1.20
## Sofia       0.08  1.49  2.00  0.26  9.90  4.25 11.02  1.90  4.77 10.90
## Stockholm   4.33  0.00 26.22  0.38  5.27  9.10  0.02  0.45  1.12  0.44
##             PC 11 PC 12
## Amsterdam    8.22  1.30
## Athens       2.84  0.08
## Berlin       3.61  0.02
## Brussels    11.80  0.42
## Budapest     1.82 16.85
## Copenhagen   4.22 18.28
## Dublin       0.00  0.01
## Elsinki      0.01  0.02
## Kiev         4.05  1.02
## Krakow       5.77 10.89
## Lisbon       8.87  2.88
## London       6.93  0.01
## Madrid       0.01  1.76
## Minsk        1.43  3.24
## Moscow       1.68  1.58
```

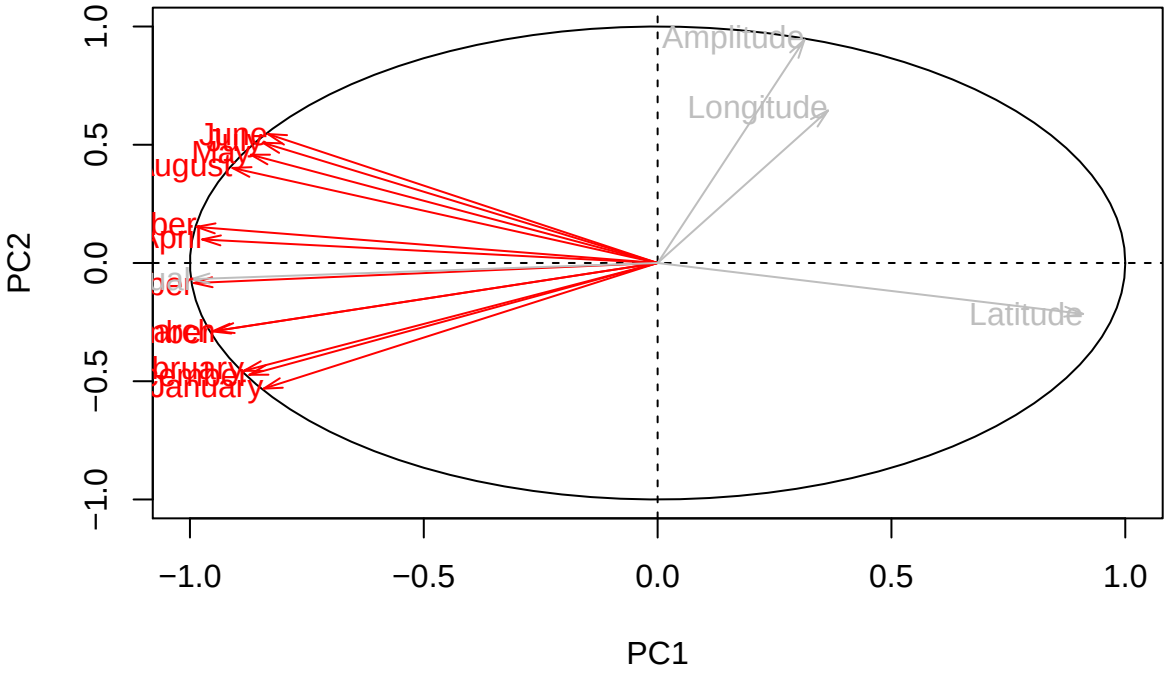
```
## Oslo        8.63 16.98
## Paris       4.18  0.75
## Prague     11.56 11.12
## Reykjavik   1.30  0.21
## Rome        4.53  0.20
## Sarajevo    8.08  0.06
## Sofia       0.00 12.33
## Stockholm   0.47  0.01
```

Contributing to
PC1: Athens, Lisbon, Rome
PC2: Reykjavik, Dublin, Moscow

4a Correlation of all

```
## [1] -0.8424506 -0.8842848 -0.9450521 -0.9738876
```

4b correlation graph



4c

Most of the active components are negativedly correlated with PC1, half positive with PC2 negative with PC2. While the sub varibles are positively correlated with PC1 except 1.

Final Conclusion

Despite not fully understanding how PCA works mathematically. From the graphs we can see it is very powerful in data compression or feature reduction.I am surprised how well the pc1, pc2 graph clusters cities

from different regions.