ExposureDA

read-in DailyPM data

```
DailyPM_Date = readRDS("/Volumes/My Passport for Mac/WD passport/Columbia-Ghana Proje
ct/MicroPEM_Data/DailyPM.rds")
```

read-in MciroPEM log data

```
require(readstata13)
```

```
## Loading required package: readstata13
```

```
## Warning: package 'readstata13' was built under R version 3.2.5
```

```
require(stringr)
```

```
## Loading required package: stringr
```

```
require(lubridate)
```

```
## Loading required package: lubridate
```

```
## Warning: package 'lubridate' was built under R version 3.2.4
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
MicroPEM = read.dta13("/Volumes/My Passport for Mac/WD passport/Columbia-Ghana Projec
t/Data/Survey_Data/MicroPem.dta")
#add leading 0 to time variables if hour is a single digit (e.g. 825 to 0825)
MicroPEM$labsetdtt = str_pad(MicroPEM$labsetdtt, 4, pad = "0")
MicroPEM$fieldsetdt = str_pad(MicroPEM$fieldsetdt, 4, pad = "0")
MicroPEM$thepaon1 = str_pad(MicroPEM$thepaon1, 4, pad = "0")
MicroPEM$pickupdtt = str_pad(MicroPEM$pickupdtt, 4, pad = "0")
MicroPEM$thepaon2 = str_pad(MicroPEM$thepaon2, 4, pad = "0")
MicroPEM$thepaoff2 = str_pad(MicroPEM$thepaoff2, 4, pad = "0")
MicroPEM$tupemoff = str_pad(MicroPEM$tupemoff, 4, pad = "0")

#assign NA to HEPA end times if Micorpem was not running when retrieving the instrume
nt
MicroPEM$thepaon2[MicroPEM$thepaon2=="0000"|MicroPEM$thepaon2=="9999"]=NA
MicroPEM$thepaoff2[MicroPEM$thepaoff2=="0000"|MicroPEM$thepaoff2=="9999"]=NA

#HEPA start Datetime
MicroPEM$HEPA1St = paste(MicroPEM$datevisit, MicroPEM$thepaon1)
MicroPEM$HEPA1St = dmy_hm(as.character(MicroPEM$HEPA1St), tz="GMT")
range(MicroPEM$HEPA1St)
```

```
## [1] "2013-01-03 10:08:00 GMT" "2016-02-24 14:31:00 GMT"
```

```
#HEPA end Datetime
MicroPEM$HEPA2St = paste(MicroPEM$pickupdtd, MicroPEM$thepaon2)
MicroPEM$HEPA2St = dmy_hm(as.character(MicroPEM$HEPA2St), tz="GMT")
```

```
## Warning: 497 failed to parse.
```

```
range(MicroPEM$HEPA2St, na.rm=T)
```

```
## [1] "2013-11-02 06:25:00 GMT" "2016-02-27 13:12:00 GMT"
```

```
MicroPEM$HEPA2End = paste(MicroPEM$pickupdtd, MicroPEM$thepaoff2)
MicroPEM$HEPA2End = dmy_hm(as.character(MicroPEM$HEPA2End), tz="GMT")
```

```
## Warning: 501 failed to parse.
```

```
range(MicroPEM$HEPA2End, na.rm=T)
```

```
## [1] "2013-11-02 06:39:00 GMT" "2016-02-27 13:19:00 GMT"
```

```
#correct a filterid typo
MicroPEM$filterid[which(is.na(MicroPEM$mstudyid))]
```

```
## character(0)
```

```
which(MicroPEM$filterid=="KHC031B")
```

```
## [1] 804
```

```
MicroPEM$filterid[MicroPEM$filterid=="KHC031B"] = "KHCD31B"
```

Merge DailyPM with MicroPEM log data

```
DailyPM = merge(DailyPM_Date, MicroPEM, by.x="filterID", by.y="filterid", all.x=T)
DailyPM[which(is.na(DailyPM$mstudyid)),]
```

```
##      filterID downloadDate totalDownloadTime deviceSerial.x
## 653   KHC0729   07-11-2014               6.6     UGF320432N
## 896   KHC1015   12-16-2014               4.4     UGF320435N
## 1885  KHCD48C     10/20/14               6.9     UGF320415N
##      dateTimeHardware dateTimeSoftware         version
## 653         15-Feb-13        09-Jul-12 v1.6.4573.15683
## 896         15-Feb-13        09-Jul-12 v1.6.4573.15683
## 1885        15-Feb-13         9-Jul-12 v1.6.4573.15683
##
participantID
## 653  /Volumes/My Passport for Mac/WD passport/Columbia-Ghana Project/MicroPem Raw
Data/Nephelometer_processed_correct/UGF320432N_KHC0729raw_BM1126M_11JUL14_s_02_pri.cs
v
## 896  /Volumes/My Passport for Mac/WD passport/Columbia-Ghana Project/MicroPem Raw
Data/Nephelometer_processed_correct/UGF320435N_KHC1015raw_BM1335M_16Dec14_s_02_pos.cs
v
## 1885 /Volumes/My Passport for Mac/WD passport/Columbia-Ghana Project/MicroPem Raw
Data/Nephelometer_processed_correct/UGF320415N_KHCD48Craw_BM1376M_20Oct14_s_02_pri.cs
v
##      participantWeight inletAerosolSize laserCyclingVariablesDelay
## 653                 NA            PM2.5                          1
## 896                 NA            PM2.5                          1
```

```
## 1885                NA            PM2.5                           1
##      laserCyclingVariablesSamplingTime laserCyclingVariablesOffTime
## 653                                  1                            8
## 896                                  1                            8
## 1885                                 1                            8
##      SystemTimes nephelometerSlope nephelometerOffset
## 653         3030                3                150
## 896         3030                3                 56
## 1885        3030                3                 46
##      nephelometerLogInterval temperatureSlope temperatureOffset
## 653                       10                1                54
## 896                       10                1                53
## 1885                      10                1                54
##      temperatureLog humiditySlope humidityOffset humidityLog
## 653              30             1             -1          10
## 896              30             1             -1          10
## 1885             30             1              0          10
##      inletPressureSlope inletPressureOffset inletPressureLog
## 653                4095                   0               30
## 896                4095                   0               30
## 1885               4095                   0               30
##      inletPressureHighTarget inletPressureLowTarget orificePressureSlope
## 653                     1280                   768                 4095
## 896                     1280                   768                 4095
## 1885                    1280                   768                 4095
##      orificePressureOffset orificePressureLog orificePressureHighTarget
## 653                      0                30                      28368
## 896                      0                30                       2100
## 1885                     0                30                       3191
##      orificePressureLowTarget flowLog flowHighTarget flowLowTarget
## 653                      1371      30            900           200
## 896                      1455      30            900           200
## 1885                     2470      30            900           200
##      flowRate accelerometerLog batteryLog ventilationSlope
## 653       0.4                5         60             <NA>
## 896       0.5                5         60             <NA>
## 1885      0.4                5         60             <NA>
##      ventilationOffset           starttime             endtime
## 653               <NA> 2014-07-07 08:11:10 2014-07-10 08:13:30
## 896               <NA> 2014-12-12 07:37:10 2014-12-14 06:00:20
## 1885              <NA> 2014-10-15 11:11:10 2014-10-18 11:51:10
##                   mintime             maxtime     mean min   max
## 653  2014-07-07 08:11:10 2014-07-10 08:13:30 31.46292  -5 10746
## 896  2014-12-12 07:37:10 2014-12-14 06:00:20 35.41944  -3 11889
## 1885 2014-10-15 11:11:10 2014-10-18 11:51:10 15.23846  -3  9186
##      startbutton button1 button2 lowbattery deadbattery timeerror
## 653            0       0       0          0           1         0
```

```
## 896            0       0       0          1           0         0
## 1885           0       0       0          0           0         0
##            starttime_new         endtime_new deviceSerial.y
## 653  2014-07-07 08:11:10 2014-07-10 08:13:30     UGF320432N
## 896  2014-12-12 07:37:10 2014-12-14 06:00:20     UGF320435N
## 1885 2014-10-15 11:11:10 2014-10-18 11:51:10     UGF320415N
##              HEPAsttime1         HEPAsttime2        HEPAendtime1
## 653  2014-07-07 08:12:10 2014-07-07 08:19:30 2014-07-10 08:10:30
## 896  2014-12-12 07:38:10 2014-12-12 07:43:30                <NA>
## 1885 2014-10-15 11:12:10 2014-10-15 11:17:20 2014-10-18 11:45:20
##              HEPAendtime2  Startdate HEPAstnumber HEPAendnumber    HEPASt
## 653  2014-07-10 08:13:30 2014-07-07           24            10 -2.909091
## 896                 <NA> 2014-12-12           18            NA  0.000000
## 1885 2014-10-18 11:50:10 2014-10-15           17            15 -2.000000
##        HEPAEnd Duration nephelometer_avg nephelometer_corr_avg     vol
## 653  -5.000000 72.05000         31.47066              35.42545 778.140
## 896         NA 46.40000         35.42966              35.42966 690.575
## 1885 -1.384615 72.68333         15.23962              16.93181 853.865
##       flow.avg     flow.sd flow.min flow.max flow28.good flow30.good
## 653  0.3600000 0.000000000     0.36     0.36           1           1
## 896  0.4961027 0.005191781     0.48     0.51           1           1
## 1885 0.3915914 0.003720629     0.37     0.41           1           1
##      Negative1 Negative2 Validity Note Harmattan netmass index
## 653          0         0        1              0  0.0461  GOOD
## 896          0         0        1              0  0.0336  GOOD
## 1885         0         0        1              0      NA  <NA>
##      duration_index flow_index       PM       CF CF_index   CF_new
## 653               1          1 52.81826 1.490969        1 1.490969
## 896               1          1 41.41476 1.168929        1 1.168929
## 1885              1          1       NA       NA        0 1.422170
##      nephelometer_final_avg             PMday_1             PMday_2
## 653                52.81826 2014-07-07 08:11:00 2014-07-08 08:11:00
## 896                41.41476 2014-12-12 07:37:00 2014-12-13 07:37:00
## 1885               24.07992 2014-10-15 11:11:00 2014-10-16 11:11:00
##                   PMday_3              PMday4 Day_1 Day_2 Day_3 OldPM_1
## 653  2014-07-09 08:11:00 2014-07-10 08:11:00     1     2     3  30.657
## 896  2014-12-14 07:37:00 2014-12-15 07:37:00     1     2     3  48.495
## 1885 2014-10-17 11:11:00 2014-10-18 11:11:00     1     2     3  20.314
##      OldPM_2 OldPM_3 CorPM_1 CorPM_2 CorPM_3 PMn_1 PMn_2 PMn_3
## 653   42.803  32.890  45.709  63.818  49.037  1440  1440  1440
## 896   21.431     NaN  56.687  25.052     NaN  1440  1344     0
## 1885  16.489  14.244  28.889  23.450  20.257  1440  1440  1440
##      compliance_1 compliance_2 compliance_3 complianceWake_1
## 653           841          879          262              776
## 896          1047          663            0              866
## 1885          730          734          776              717
##      complianceWake_2 complianceWake_3 PMAverage24 PMAverage48 PMAverage72
```

```
## 653              796              199      45.709      54.764      52.855
## 896              645                0      56.687      41.415      41.415
## 1885             701              725      28.889      26.169      24.199
##      batchno formno mstudyid mpermid compno vname cluster datevisit   ls
## 653       NA     NA     <NA>    <NA>   <NA>  <NA>    <NA>      <NA> <NA>
## 896       NA     NA     <NA>    <NA>   <NA>  <NA>    <NA>      <NA> <NA>
## 1885      NA     NA     <NA>    <NA>   <NA>  <NA>    <NA>      <NA> <NA>
##       ftd vround labsetdtd labsetdtt upemid flowca nephzero nephvalue
## 653  <NA>   <NA>      <NA>      <NA>   <NA>   <NA>     <NA>        NA
## 896  <NA>   <NA>      <NA>      <NA>   <NA>   <NA>     <NA>        NA
## 1885 <NA>   <NA>      <NA>      <NA>   <NA>   <NA>     <NA>        NA
##      nephslope lasapara activinter samponoff fieldsetd fieldsetdt thepaon1
## 653         NA     <NA>       <NA>      <NA>      <NA>       <NA>     <NA>
## 896         NA     <NA>       <NA>      <NA>      <NA>       <NA>     <NA>
## 1885        NA     <NA>       <NA>      <NA>      <NA>       <NA>     <NA>
##      thepaoff1 pickupdtd pickupdtt upemun ncomment inletconne nocomment
## 653         NA      <NA>      <NA>   <NA>     <NA>       <NA>      <NA>
## 896         NA      <NA>      <NA>   <NA>     <NA>       <NA>      <NA>
## 1885        NA      <NA>      <NA>   <NA>     <NA>       <NA>      <NA>
##      thepaon2 thepaoff2 tupemoff computer dbackup initialsba comments
## 653      <NA>      <NA>     <NA>     <NA>    <NA>       <NA>     <NA>
## 896      <NA>      <NA>     <NA>     <NA>    <NA>       <NA>     <NA>
## 1885     <NA>      <NA>     <NA>     <NA>    <NA>       <NA>     <NA>
##      HEPA1St HEPA2St HEPA2End
## 653     <NA>    <NA>     <NA>
## 896     <NA>    <NA>     <NA>
## 1885    <NA>    <NA>     <NA>
```

missing three MircoPEM logsheet KHC0729, KHC1015, KHCD48C

read-in DailyCO data

```
COdata = readRDS("/Users/zhengzhou/Dropbox/Ghana_exposure_data_SHARED_2014/CO_files_p
rocessed/FINAL_CO_parameters_withvalidation_2016Jun14.rds")
COdata$Startdate = as.Date(COdata$firstdate)                                    #get the
start date of CO measurements
COdata1 = COdata[is.na(COdata$cstudyid),]                                       #exclude
child CO measurements
```

Merge PM and CO data

```
PMCO = merge(DailyPM, COdata1, by=c("mstudyid", "Startdate"), all.x=T)
```

Wide to long format

```
# change co colnames
colnames(PMCO)[colnames(PMCO)=="co_day1_mean"] = "OldCO_1"
colnames(PMCO)[colnames(PMCO)=="co_day2_mean"] = "OldCO_2"
colnames(PMCO)[colnames(PMCO)=="co_day3_mean"] = "OldCO_3"
colnames(PMCO)[colnames(PMCO)=="co_day1_mean_corr"] = "CorCO_1"
colnames(PMCO)[colnames(PMCO)=="co_day2_mean_corr"] = "CorCO_2"
colnames(PMCO)[colnames(PMCO)=="co_day3_mean_corr"] = "CorCO_3"

PMCO1 <-reshape(PMCO,
                    varying=c(grep("PMday_",colnames(PMCO)),
                              grep("OldPM_",colnames(PMCO)),
                              grep("CorPM_",colnames(PMCO)),
                              grep("compliance_",colnames(PMCO)),
                              grep("complianceWake_",colnames(PMCO)),
                              grep("PMn_",colnames(PMCO)),
                              grep("OldCO_",colnames(PMCO)),
                              grep("CorCO_",colnames(PMCO)),
                              grep("Day_",colnames(PMCO))),
                    idvar="id",
                    direction="long",  sep="_")
```

Data cleaning of the merged PMCO data

```
PMCO2 = PMCO1[PMCO1$visually_valid!=3,]    # exclude samples with invalid CO readings
PMCO2 = PMCO2[PMCO2$PMn>1320,]      #exclude PM sample-day < 22hrs
PMCO2 = PMCO2[!is.na(PMCO2$CorCO),]  #exclude CO sample-day < 24hrs
summary(PMCO2$CorPM)       # check the range of PM
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -13.27   39.35   62.81   78.54  100.40  610.80
```

```
summary(PMCO2$CorCO)       # check the range of CO
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.1260  0.4873  1.0530  1.2710 91.7100
```

```
PMCO2 = PMCO2[PMCO2$CorPM>0,]       #exclude PM <0
#calculate compliance measure and categorize the measure into 7 buckets
PMCO2$complianceWakePct = PMCO2$complianceWake/PMCO2$PMn
PMCO2$complianceWakePctGP = cut(PMCO2$complianceWakePct, seq(0, 0.7, 0.1), labels=c(0
.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7), right=FALSE)
```

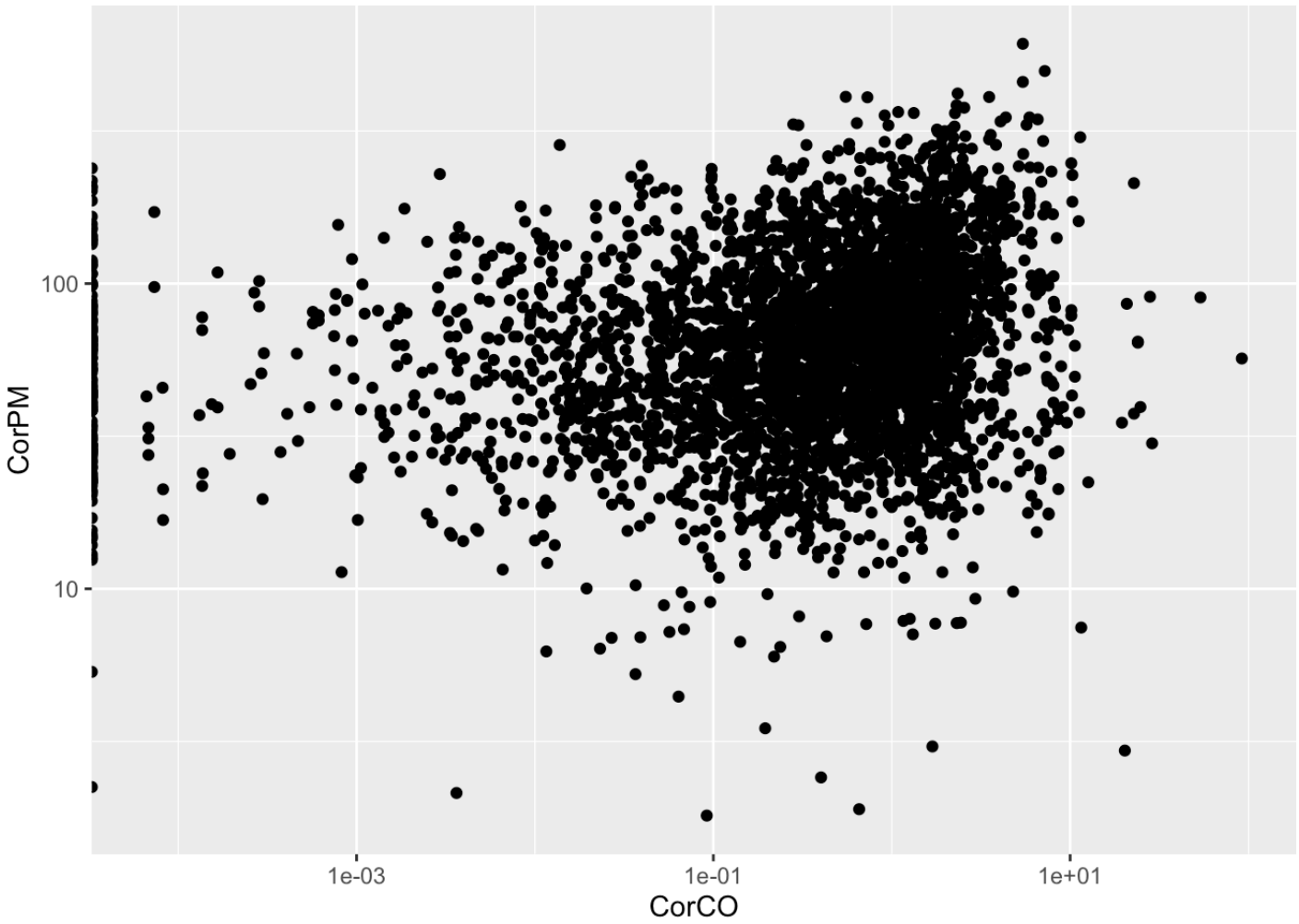
Plot PM vs CO

```
require(ggplot2)
```

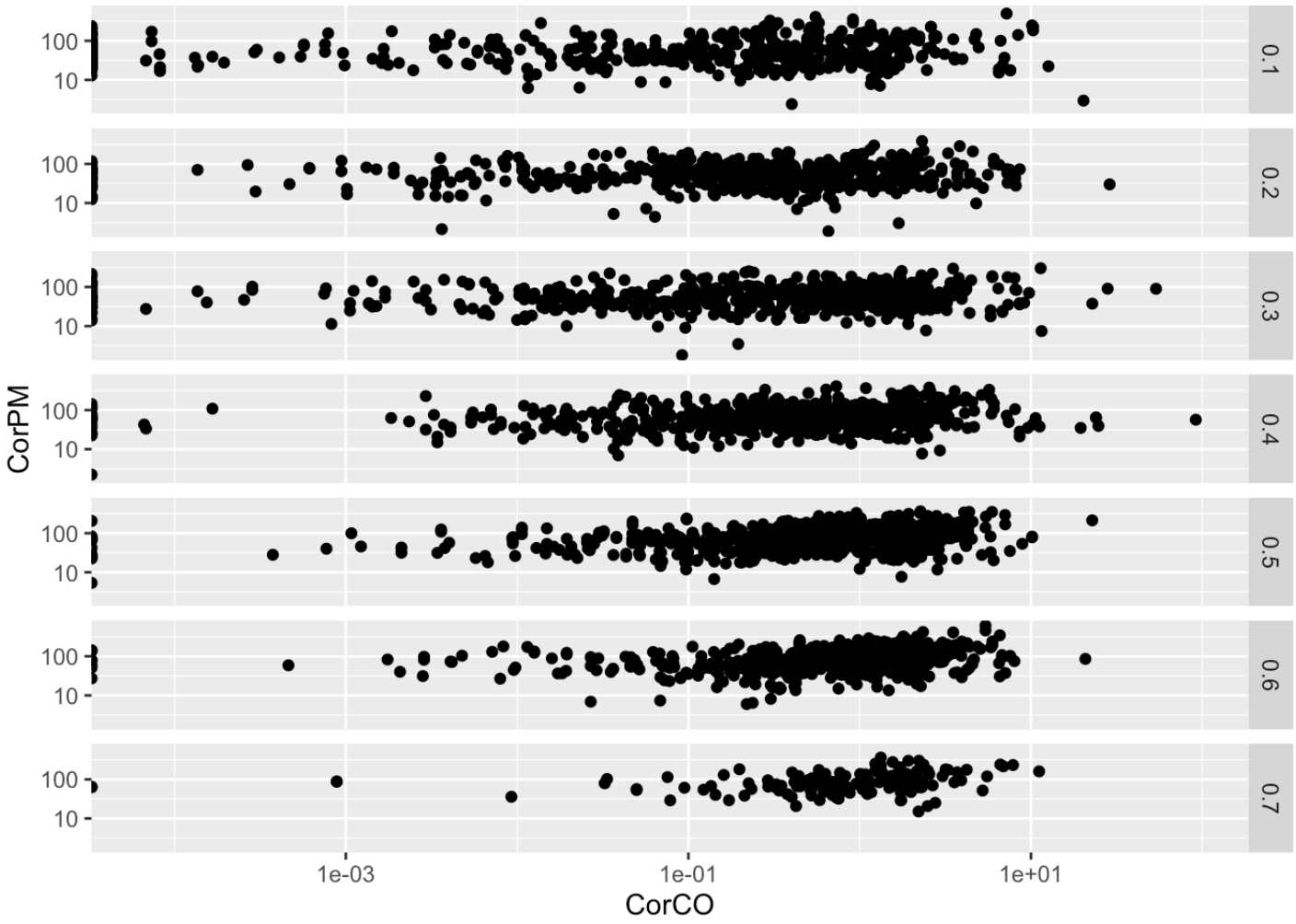
```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

```
par(mar=c(4,4,1,1))
par(las=3)
ggplot(PMCO2, aes(CorCO, CorPM)) + geom_point() + scale_x_log10() + scale_y_log10()
```



Plot PM vs CO by compliance bin



Correlation between PM and CO by compliance bin

```
cor(PMCO2$CorCO[PMCO2$complianceWakePctGP==0.1], PMCO2$CorPM[PMCO2$complianceWakePctGP==0.1])
```

```
## [1] 0.1292247
```

```
cor(PMCO2$CorCO[PMCO2$complianceWakePctGP==0.2], PMCO2$CorPM[PMCO2$complianceWakePctGP==0.2])
```

```
## [1] 0.04321132
```

```
cor(PMCO2$CorCO[PMCO2$complianceWakePctGP==0.3], PMCO2$CorPM[PMCO2$complianceWakePctGP==0.3])
```

```
## [1] 0.08690393
```

```
cor(PMCO2$CorCO[PMCO2$complianceWakePctGP==0.4], PMCO2$CorPM[PMCO2$complianceWakePctGP==0.4])
```

```
## [1] 0.03021688
```

```
cor(PMCO2$CorCO[PMCO2$complianceWakePctGP==0.5], PMCO2$CorPM[PMCO2$complianceWakePctGP==0.5])
```

```
## [1] 0.2767232
```

```
cor(PMCO2$CorCO[PMCO2$complianceWakePctGP==0.6], PMCO2$CorPM[PMCO2$complianceWakePctGP==0.6])
```

```
## [1] 0.2839204
```

```
cor(PMCO2$CorCO[PMCO2$complianceWakePctGP==0.7], PMCO2$CorPM[PMCO2$complianceWakePctGP==0.7])
```

```
## [1] 0.3808079
```

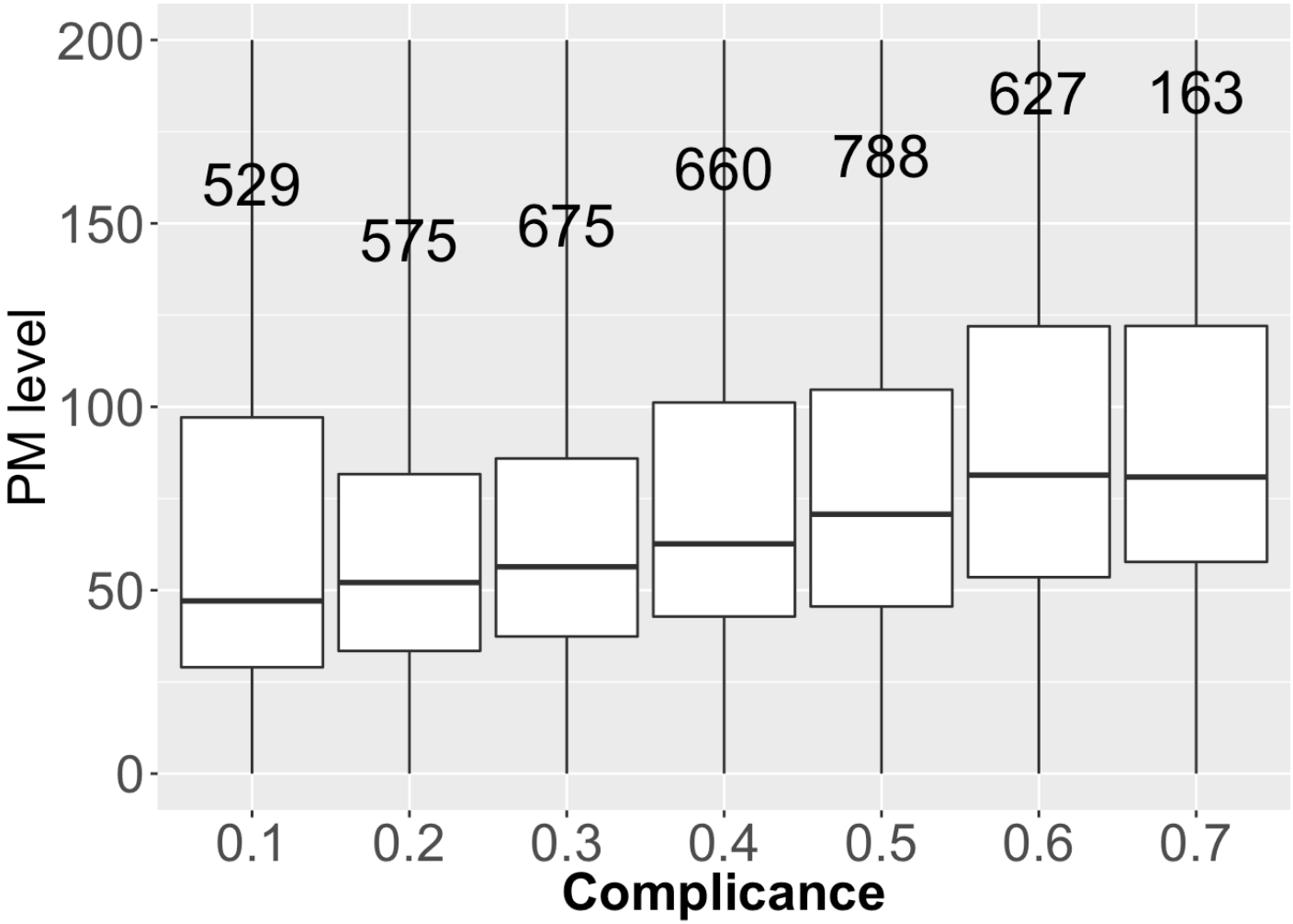
Boxplot of PM by compliance bin

```
## Loading required package: plyr
```

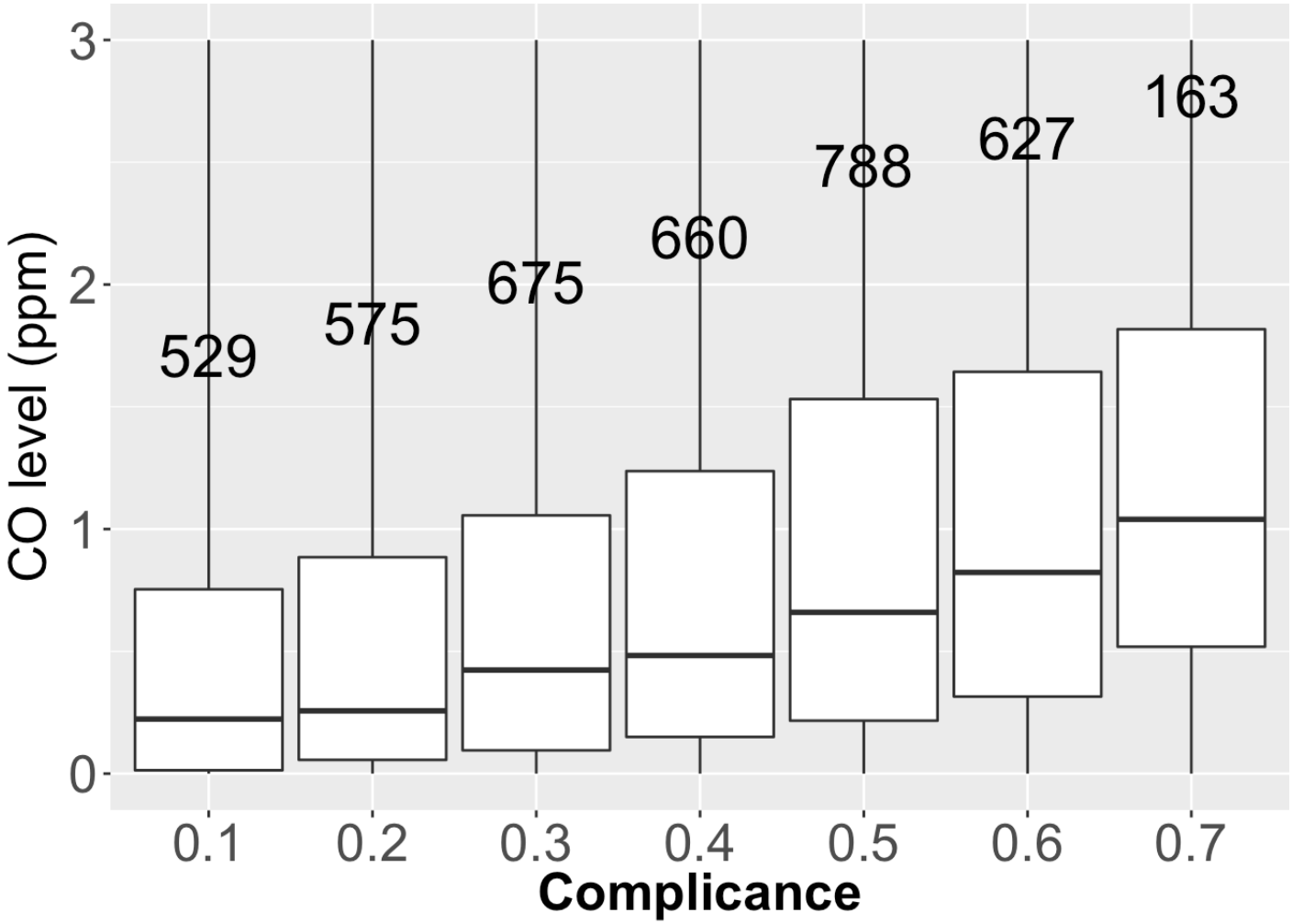
```
## Warning: package 'plyr' was built under R version 3.2.5
```

```
##
## Attaching package: 'plyr'
```

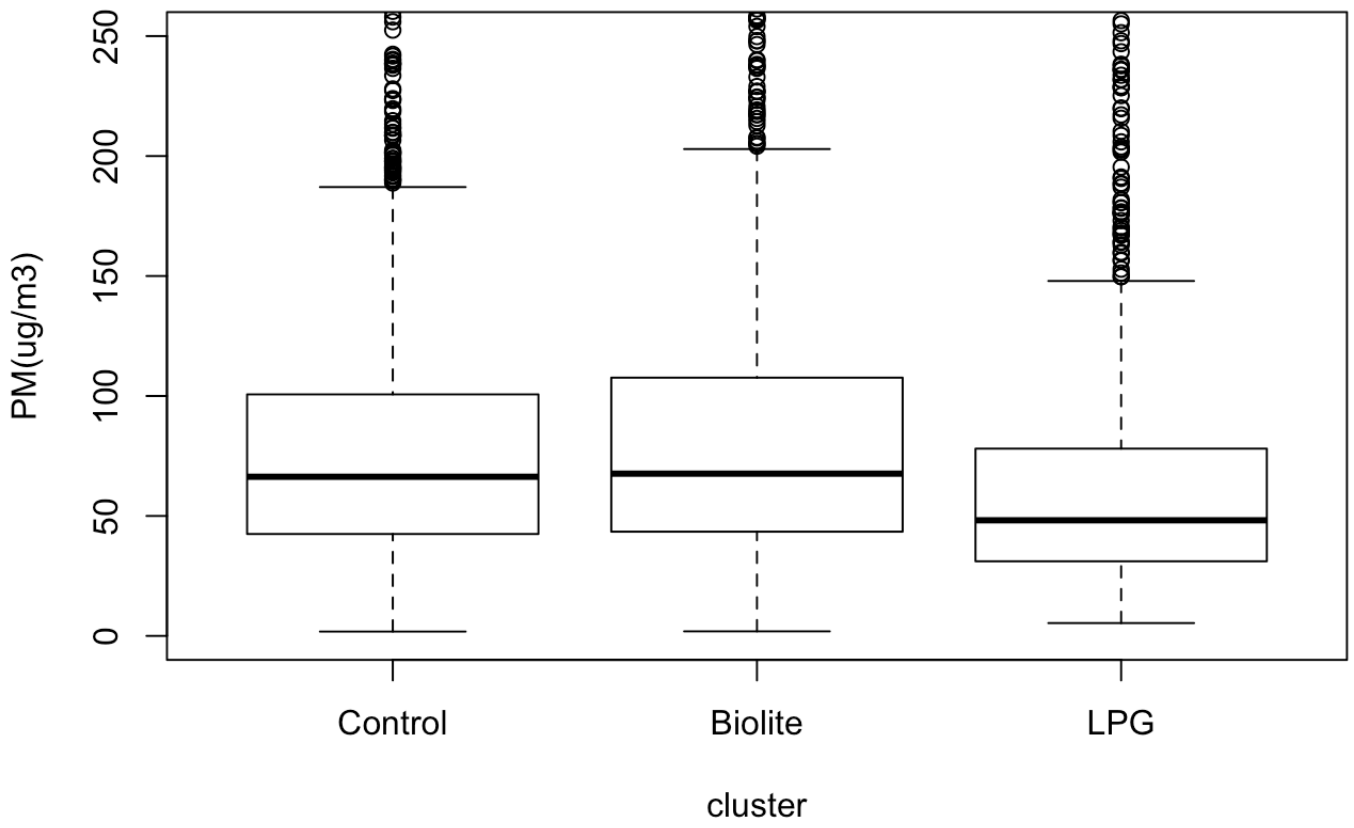
```
## The following object is masked from 'package:lubridate':
##
##     here
```



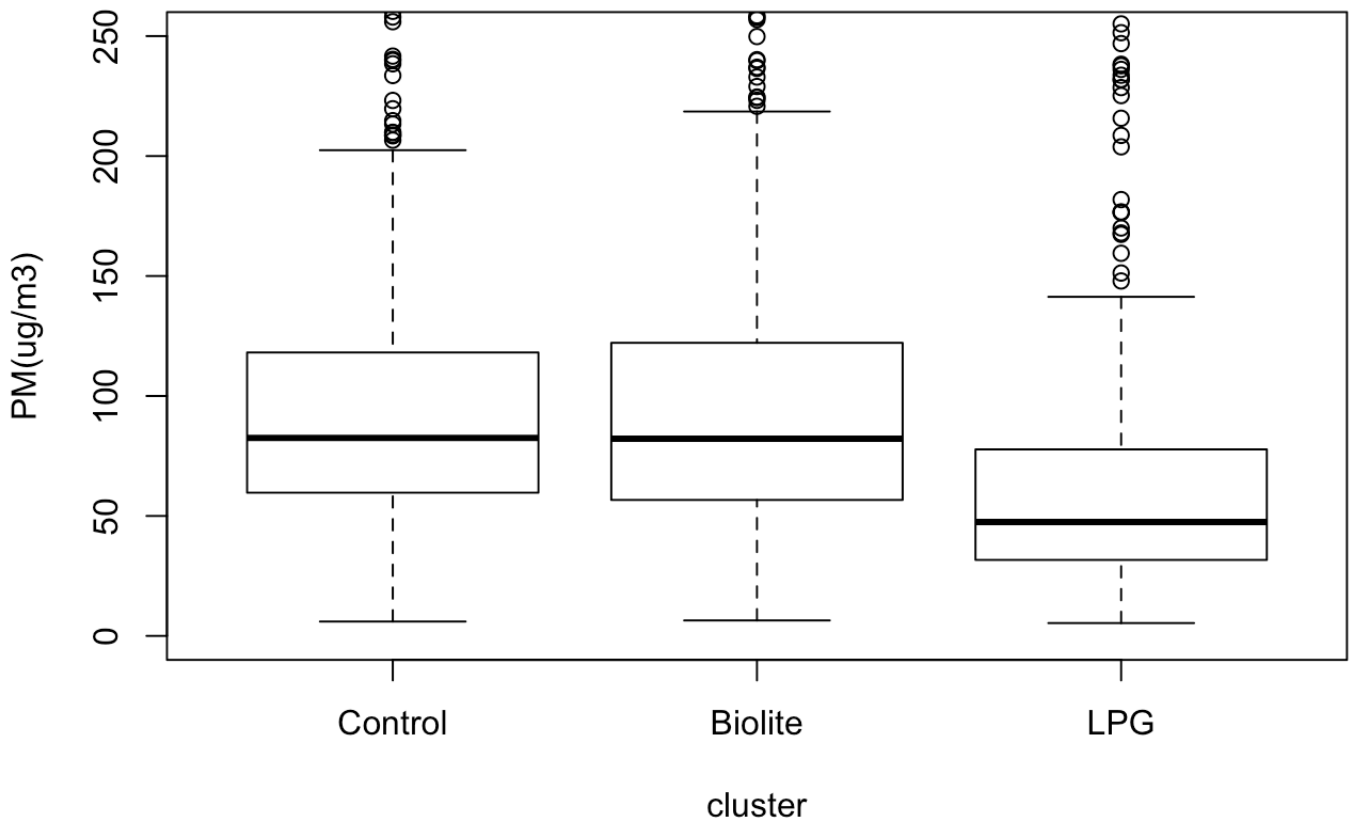
Boxplot of CO by compliance bin



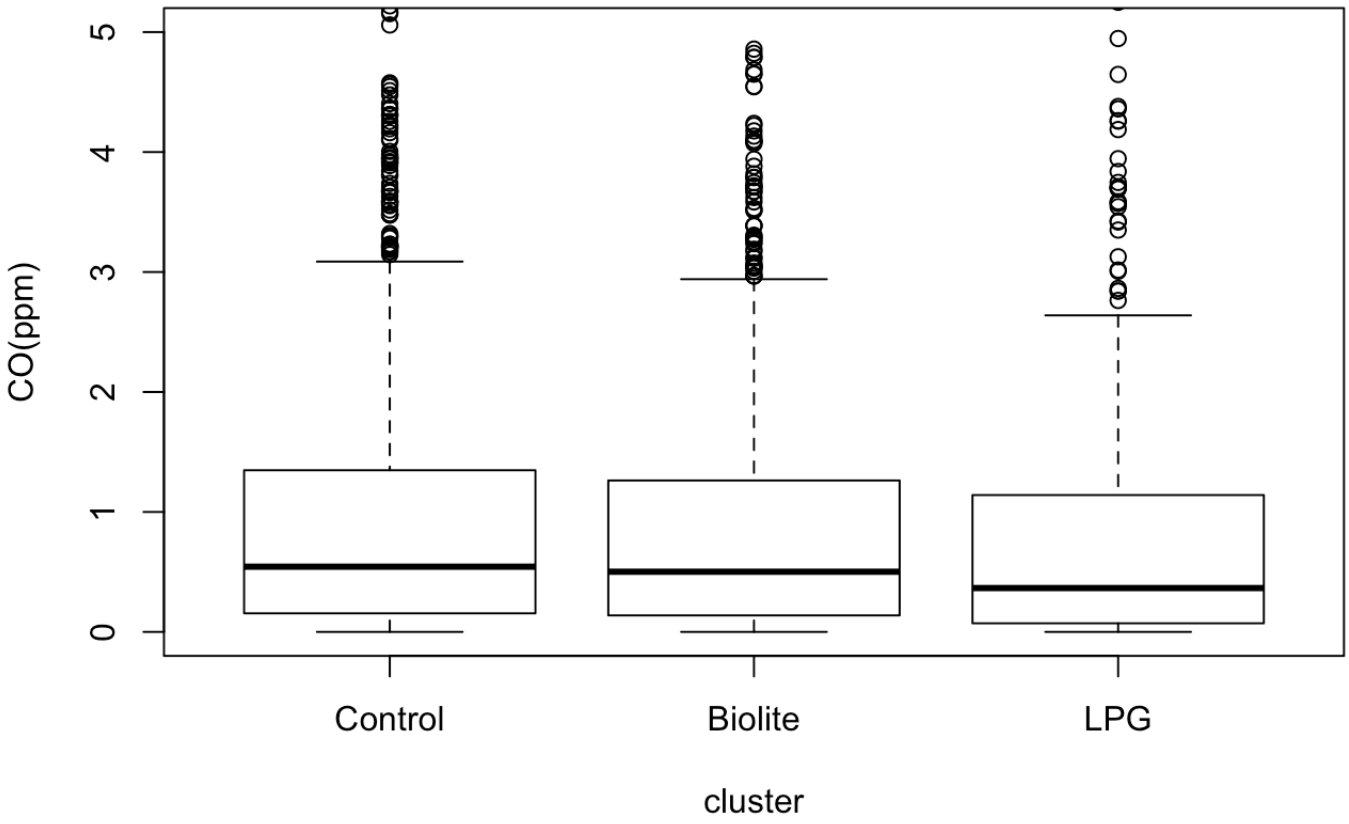
Boxplot of PM by arm



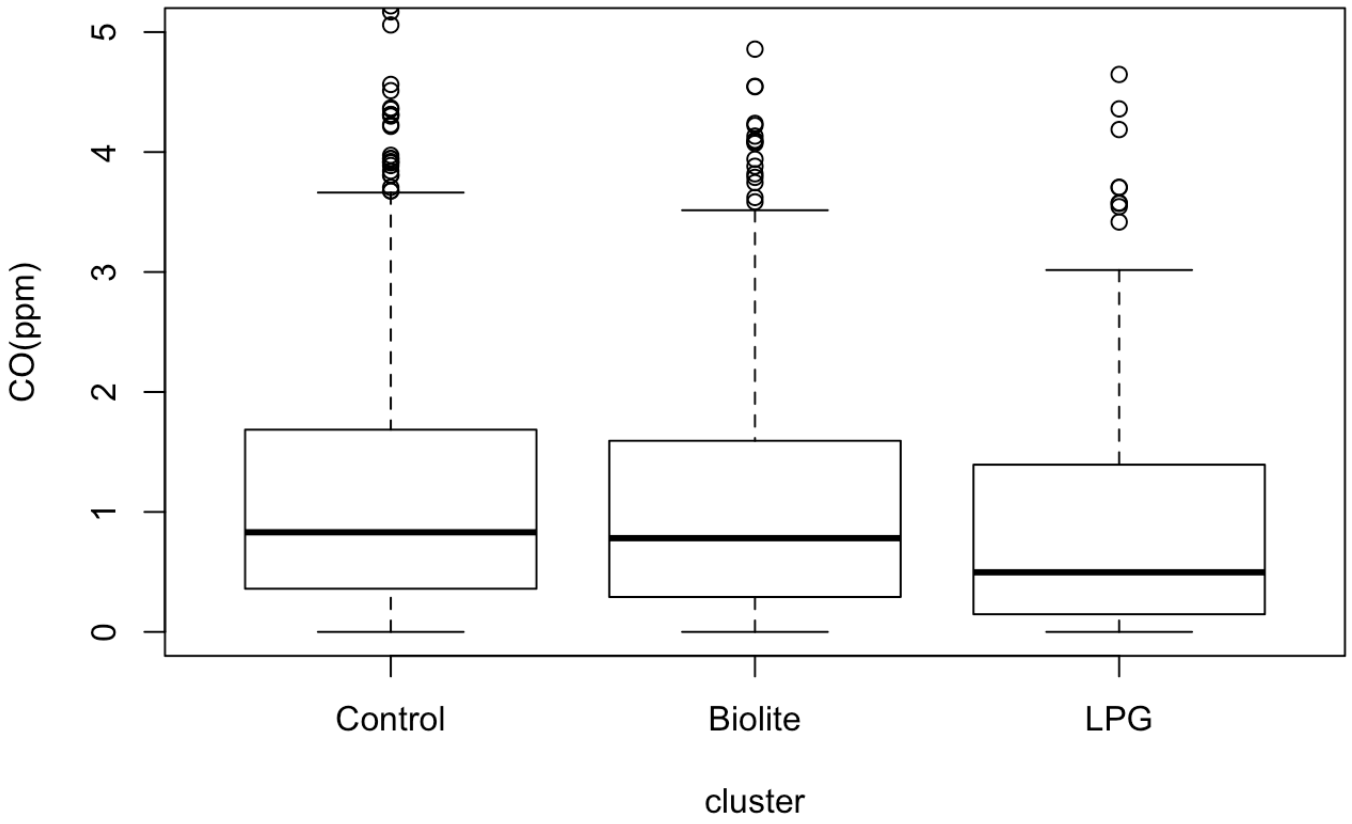
Boxplot of PM by arm (compliance > 40%)



Boxplot of CO by arm



Boxplot of CO by arm (compliance > 40%)



read in daily exposure data

```
Exposure = read.dta13("/Volumes/My Passport for Mac/WD passport/Columbia-Ghana Projec
t/Data/Survey_Data/Exposure.dta", generate.factors=T)
Exposure$Startdate = as.Date(dmy(Exposure$datevisit, tz="GMT")) -1     # create a var
oable for the startdate of daily exposure
which(duplicated(Exposure[c("mstudyid","Startdate")]))        # check duplicate daily
exposure data
```

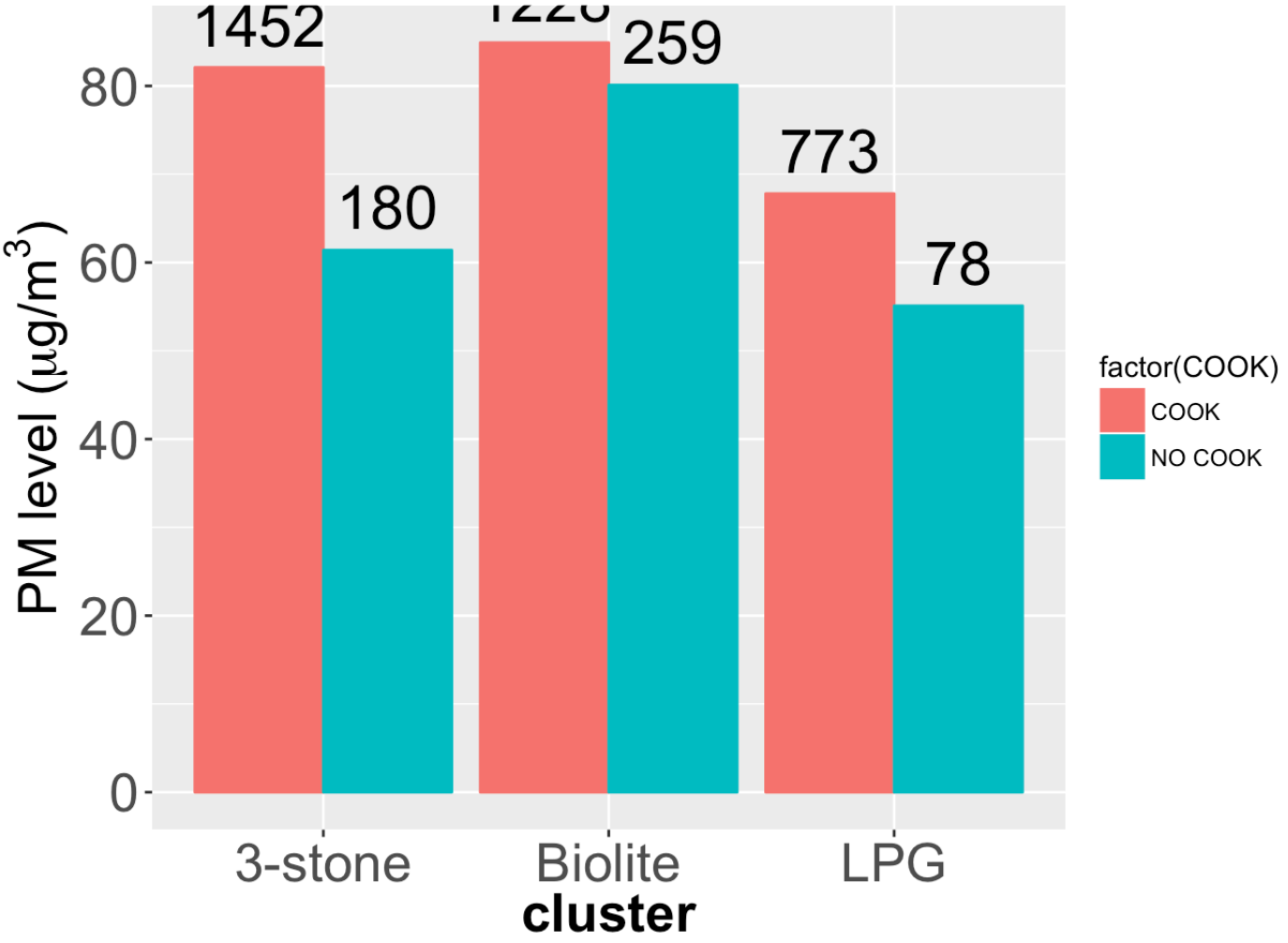
```
##  [1]  1106  1108  1111  1113  1615  1663  2331  2333  2547  3492  3496
## [12]  3501  3582  3583  3662  3663  3899  3945  4143  5159  5431  5558
## [23]  5608  6892  7066  7294  7451  7890  8177  8208  8616  8674  9225
## [34]  9492  9809  9983 10250 10332 10571 10650 10653 10906 10923 11537
## [45] 11539 12683 12839 12857 13187 14718 15061 15207 15395 15480 15612
## [56] 16078 16208 16250 16999 17001 17114 17115 17125 17126 17547 17628
## [67] 18460 18552 18650 18660 18841 18883 18986 19019 19497 19501 19611
## [78] 19839 20634 21015 21446 21726 21771 21915 22538 23531 23738 23944
## [89] 23986 24103 24173 24866 24994 24995
```

some duplicated daily exposure data

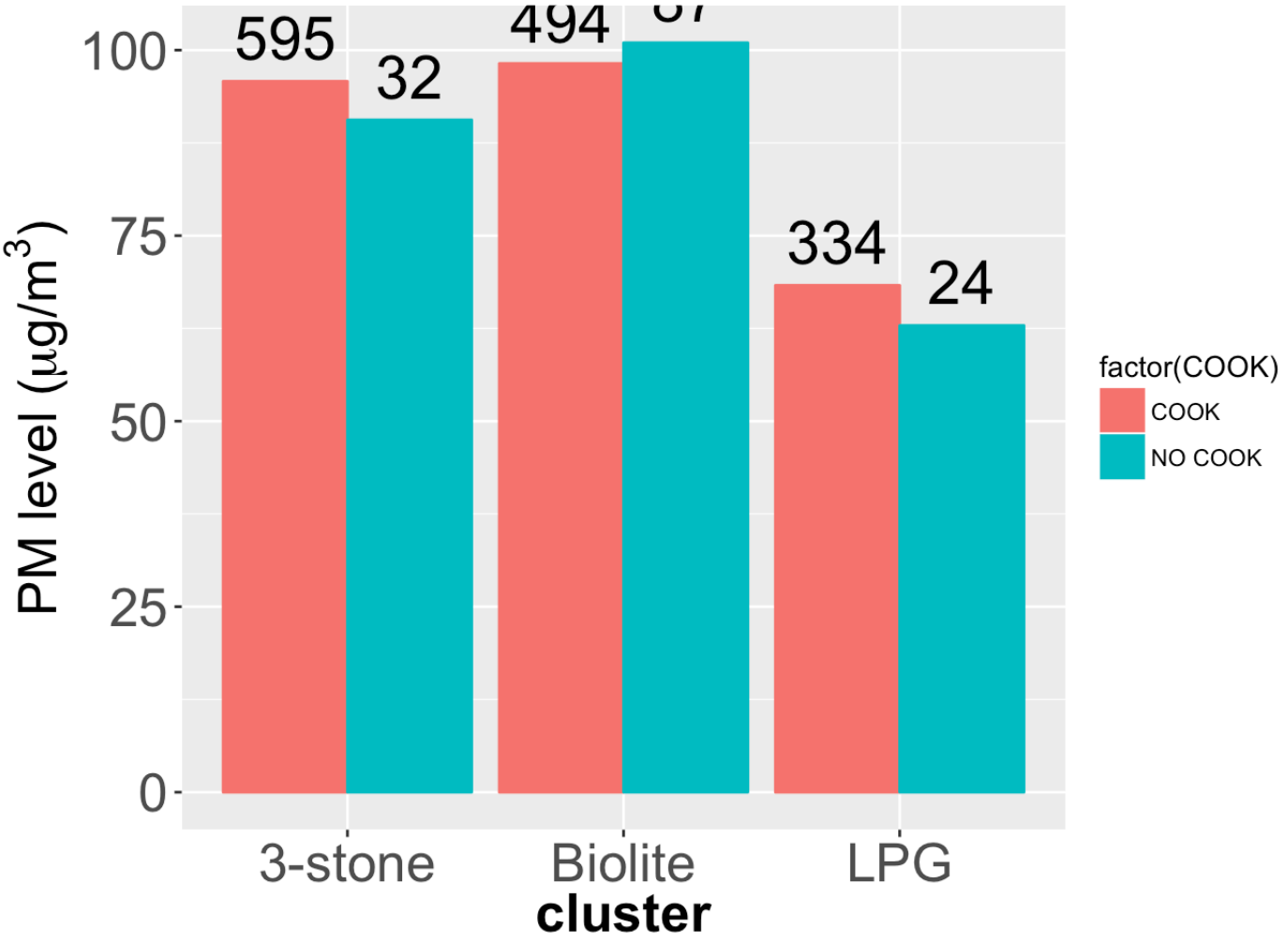
Merge PMCO with daily exposure data

```
PMCO3 = merge(PMCO2, Exposure, by = c("mstudyid","Startdate"), all.x= T )
```

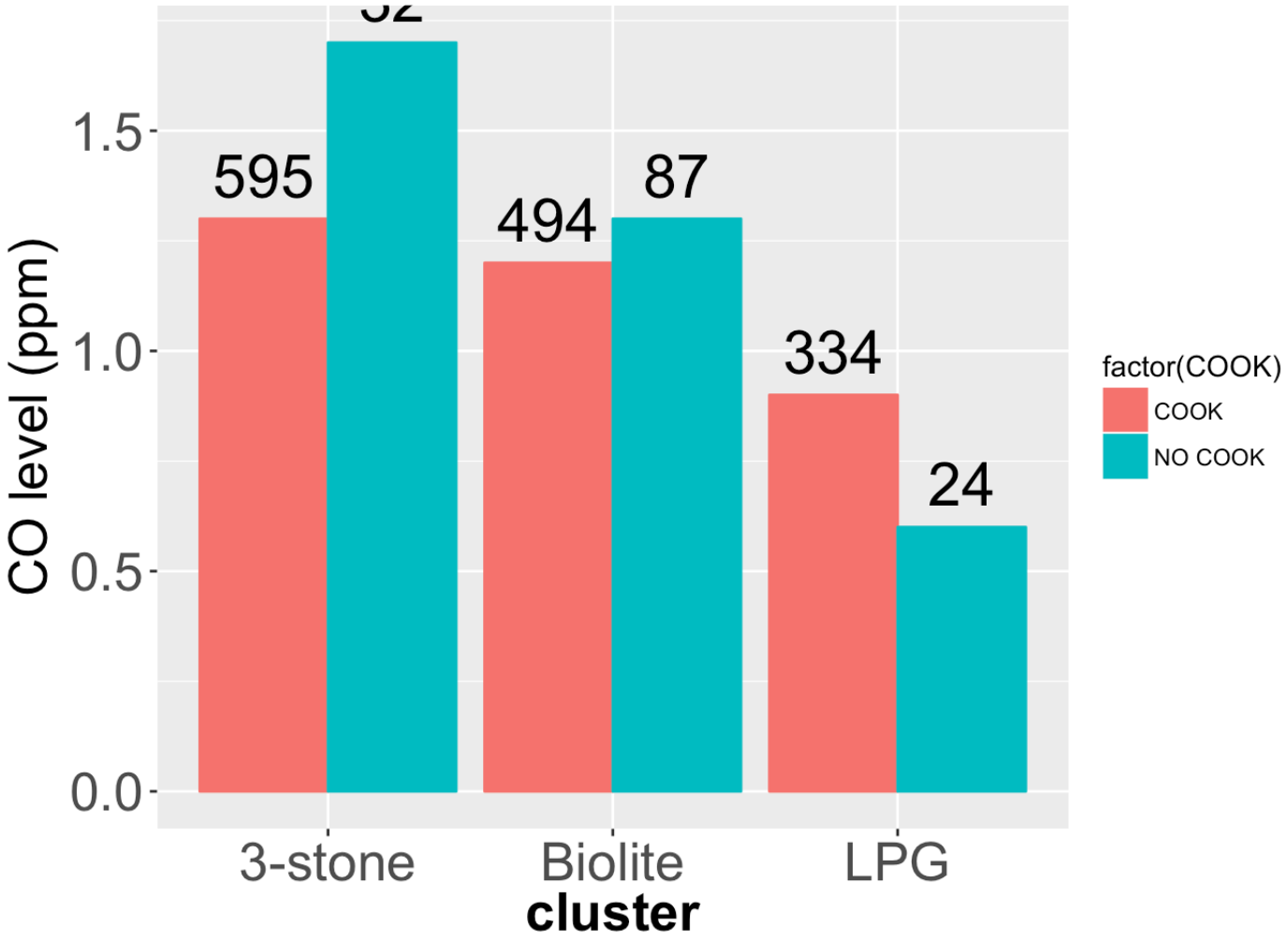
Plot PM by arm and cookfood



Plot PM by arm and cookfood (compliance > 40%)

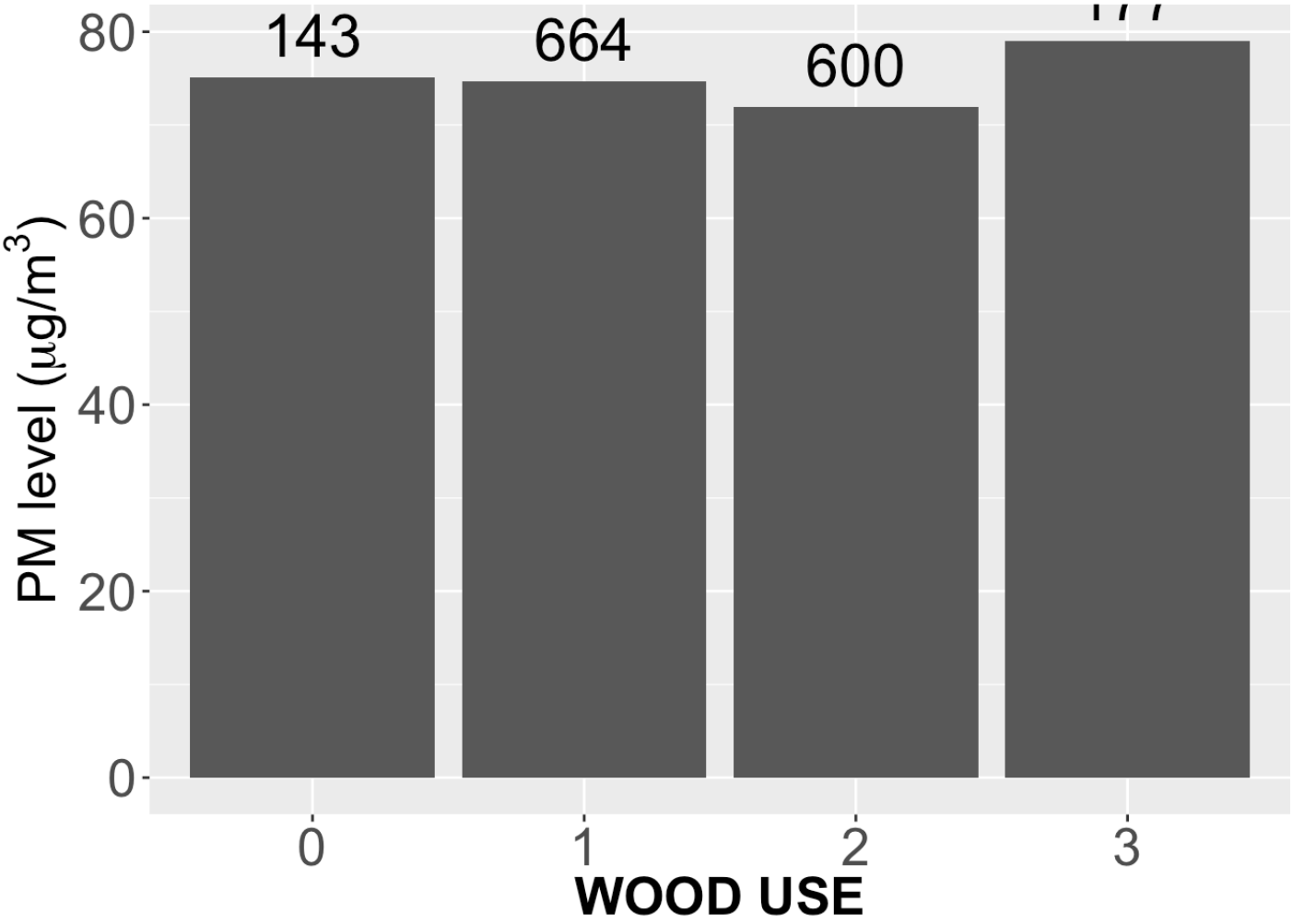


Plot CO by arm and cookfood (compliance > 40%)



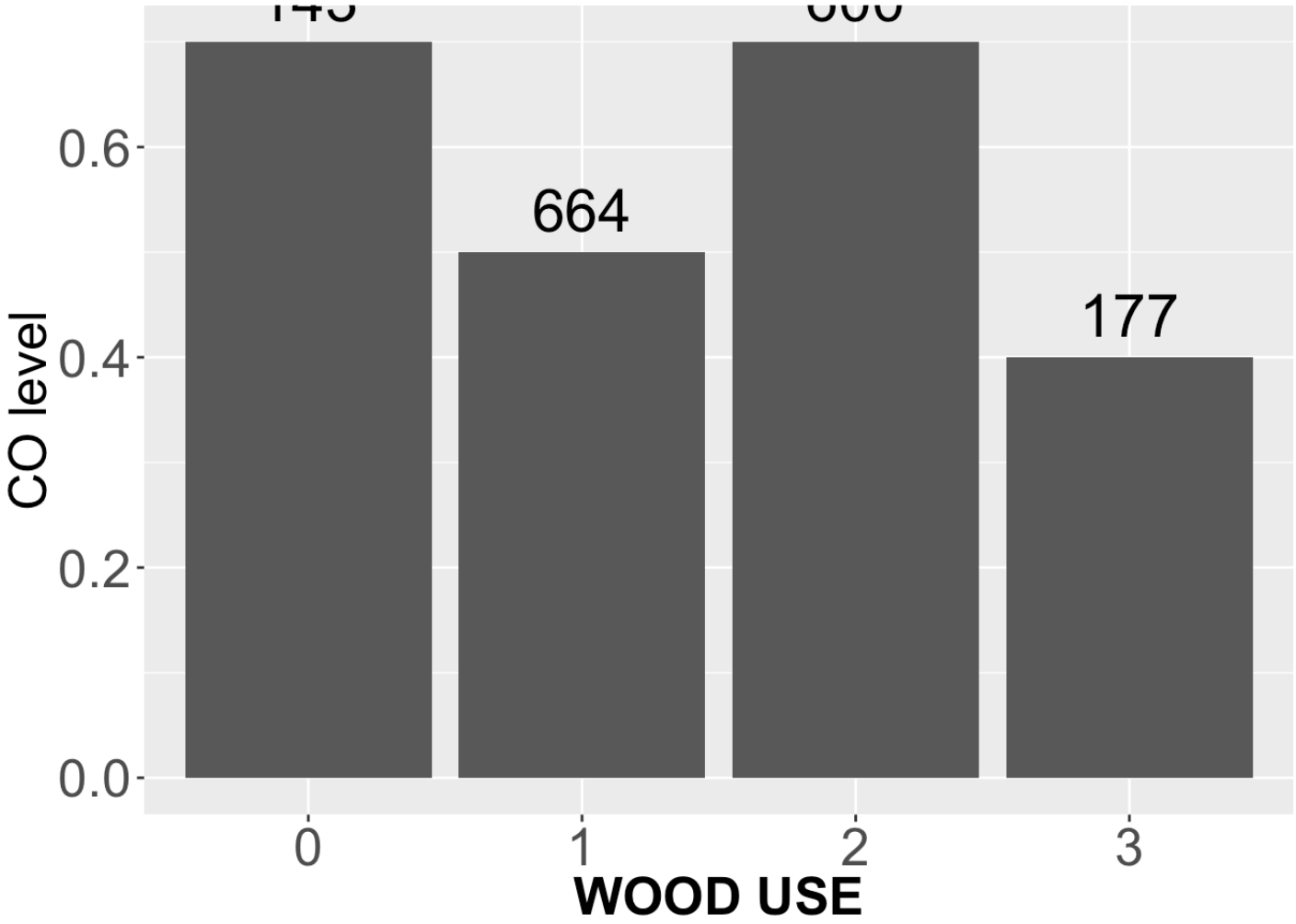
Plot PM by fuel (compliance > 40%)

```
##   WOOD     m   gm   md     min     max counts
## 1    0 92.3 75.1 79.6   6.702  419.66    143
## 2    1 90.4 74.7 75.7   6.908 610.771    664
## 3    2 88.7 71.9 72.5   5.345 362.191    600
## 4    3 94.8   79 80.5 17.231 318.879    177
```



Plot CO by fuel (compliance > 40%)

```
##   WOOD   m  gm  md min              max counts
## 1    0 1.3 0.7 0.9   0 7.99826591468949    143
## 2    1 1.1 0.5 0.7   0 22.7802579365079    664
## 3    2 1.2 0.7 0.8   0 11.1537990196078    600
## 4    3 0.9 0.4 0.6   0 6.05438643086281    177
```


0.6
0.4
0.2
0.0
CO level
664
177
0
1
2
3
WOOD USE