

## EDA\_US\_PollutionData

### I. Read in the data set

*# Data set of air pollution in US from 2000 to 2016.*

```
pollution <-  
data.frame(read.csv("pollution_data/pollution_us_2000_2016.csv"))
```

### II. Subset data and view the data frame structures

```
attach(pollution)  
names(pollution)
```

```
## [1] "X"                "State.Code"        "County.Code"  
## [4] "Site.Num"         "Address"           "State"  
## [7] "County"          "City"              "Date.Local"  
## [10] "NO2.Units"        "NO2.Mean"          "NO2.1st.Max.Value"  
## [13] "NO2.1st.Max.Hour" "NO2.AQI"           "O3.Units"  
## [16] "O3.Mean"          "O3.1st.Max.Value"  "O3.1st.Max.Hour"  
## [19] "O3.AQI"           "SO2.Units"         "SO2.Mean"  
## [22] "SO2.1st.Max.Value" "SO2.1st.Max.Hour"  "SO2.AQI"  
## [25] "CO.Units"         "CO.Mean"           "CO.1st.Max.Value"  
## [28] "CO.1st.Max.Hour"  "CO.AQI"
```

*# Subset data for MD, VA, DC, and tri (three states combined).*

```
pollution_md <- subset(pollution, State == "Maryland")[, -(1:5)]  
pollution_va <- subset(pollution, State == "Virginia")[, -(1:5)]  
pollution_dc <- subset(pollution, State == "District Of Columbia")[, -(1:5)]  
pollution_tri <- rbind(pollution_dc, pollution_md, pollution_va)  
pollution_tri <- droplevels.data.frame(pollution_tri)
```

*# Summary of the three states.*

```
summary(pollution_tri)
```

```
##              State              County  
## District Of Columbia:25696 District of Columbia:25696  
## Maryland             :23538 Fairfax             :16622  
## Virginia             :36422 Baltimore          :13174  
##                      Prince George's           : 7384  
##                      Alexandria City            : 6936  
##                      Henrico                    : 5976  
##                      (Other)                   : 9868  
##              City              Date.Local              NO2.Units  
## Washington           :25696 2013-03-31: 32 Parts per billion:85656  
## Essex                :13174 2013-04-01: 32  
## Beltsville           : 7384 2013-04-02: 32  
## Not in a city        : 6590 2013-04-03: 32  
## East Highland Park: 5976 2013-04-04: 32  
## Alexandria           : 5876 2013-04-05: 32
```

```

## (Other) :20960 (Other) :85464
## NO2.Mean NO2.1st.Max.Value NO2.1st.Max.Hour NO2.AQI
## Min. : 0.000 Min. : 0.00 Min. : 0.00 Min. : 0.00
## 1st Qu.: 6.167 1st Qu.: 14.00 1st Qu.: 6.00 1st Qu.: 13.00
## Median :11.000 Median : 24.00 Median : 8.00 Median : 23.00
## Mean :12.527 Mean : 24.78 Mean :11.64 Mean : 23.23
## 3rd Qu.:17.348 3rd Qu.: 34.00 3rd Qu.:20.00 3rd Qu.: 32.00
## Max. :65.208 Max. :141.00 Max. :23.00 Max. :109.00
##
## 03.Units 03.Mean 03.1st.Max.Value
## Parts per million:85656 Min. :0.00000 Min. :0.00000
## 1st Qu.:0.01771 1st Qu.:0.02800
## Median :0.02642 Median :0.03800
## Mean :0.02642 Mean :0.03933
## 3rd Qu.:0.03463 3rd Qu.:0.05000
## Max. :0.07362 Max. :0.12800
##
## 03.1st.Max.Hour 03.AQI S02.Units
## Min. : 0.0 Min. : 0.00 Parts per billion:85656
## 1st Qu.: 9.0 1st Qu.: 24.00
## Median :10.0 Median : 33.00
## Mean :10.2 Mean : 36.45
## 3rd Qu.:11.0 3rd Qu.: 44.00
## Max. :23.0 Max. :206.00
##
## S02.Mean S02.1st.Max.Value S02.1st.Max.Hour S02.AQI
## Min. : -1.5125 Min. : -1.400 Min. : 0.00 Min. : 0
## 1st Qu.: 0.6375 1st Qu.: 1.400 1st Qu.: 7.00 1st Qu.: 1
## Median : 1.7750 Median : 3.900 Median :11.00 Median : 6
## Mean : 2.9874 Mean : 6.344 Mean :10.93 Mean : 10
## 3rd Qu.: 4.2500 3rd Qu.: 8.600 3rd Qu.:15.00 3rd Qu.: 14
## Max. :42.9167 Max. :173.000 Max. :23.00 Max. :145
## NA's :42812
## CO.Units CO.Mean CO.1st.Max.Value
## Parts per million:85656 Min. : -0.4375 Min. : -0.4000
## 1st Qu.: 0.2125 1st Qu.: 0.3000
## Median : 0.3375 Median : 0.5000
## Mean : 0.4730 Mean : 0.6839
## 3rd Qu.: 0.6125 3rd Qu.: 0.9000
## Max. : 3.5304 Max. : 8.6000
##
## CO.1st.Max.Hour CO.AQI
## Min. : 0.000 Min. : 0.00
## 1st Qu.: 0.000 1st Qu.: 3.00
## Median : 6.000 Median : 5.00
## Mean : 8.193 Mean : 6.96
## 3rd Qu.:16.000 3rd Qu.: 9.00
## Max. :23.000 Max. :58.00
## NA's :42820

```

### III. Descriptive statistics

We decided to analyze Air Quality Index (AQI) of each of the 4 pollutants in Maryland, Virginia, and Washington DC for this project. The pollutants include: - Nitrogen Dioxide (NO2) - Ozone (O3) - Sulfur Dioxide (SO2) - Carbon Monoxide (CO)

Descriptive statistics were included to give us a better sense of the variables we are working with.

```
library(pastecs)

## Warning: package 'pastecs' was built under R version 3.4.3

## Loading required package: boot

## Warning: package 'boot' was built under R version 3.4.3

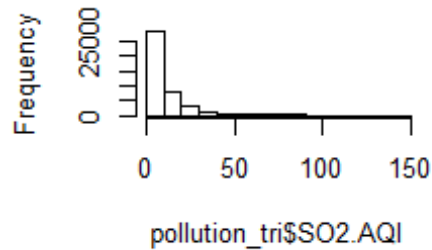
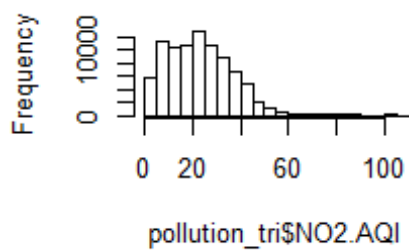
# Descriptive statistics on pollutant AQIs for the 3 states.
desc_tri <- stat.desc(pollution_tri, options(scipen = 999))
desc_triAQI <- round(desc_tri[, c("NO2.AQI", "O3.AQI", "SO2.AQI", "CO.AQI")],
  digits = 2)
desc_triAQI

##           NO2.AQI O3.AQI SO2.AQI CO.AQI
## median      23.00  33.00   6.00   5.00
## mean        23.23  36.45  10.00   6.96
## SE.mean      0.04   0.07   0.06   0.03
## CI.mean      0.09   0.13   0.12   0.06
## var         172.60 402.43 151.62  35.78
## std.dev      13.14 20.06  12.31   5.98
## coef.var     0.57  0.55   1.23   0.86

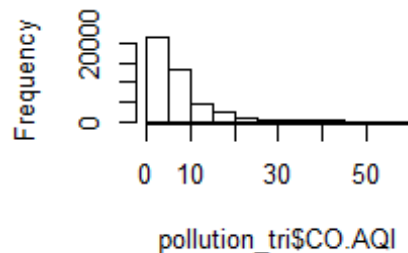
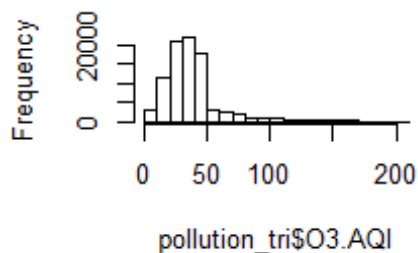
options(scipen = 0)

# Histograms for AQI of each of the 4 pollutants in MD, VA, and DC, to show
how the data is distributed.
par(mfrow = c(2, 2))
hist(pollution_tri$NO2.AQI)
hist(pollution_tri$SO2.AQI)
hist(pollution_tri$O3.AQI)
hist(pollution_tri$CO.AQI)
```

## Histogram of pollution\_tri\$NO2. Histogram of pollution\_tri\$SO2.



## Histogram of pollution\_tri\$O3. Histogram of pollution\_tri\$CO.



### IV. Boxplots

Using data transformation techniques, we drilled down to the county level for MD and VA using boxplots to see how the pollutant AQIs are distributed among counties with varying populations. Note that we used the same analysis for DC, even though it has only one county.

#### IV(a). Boxplot for counties in Maryland

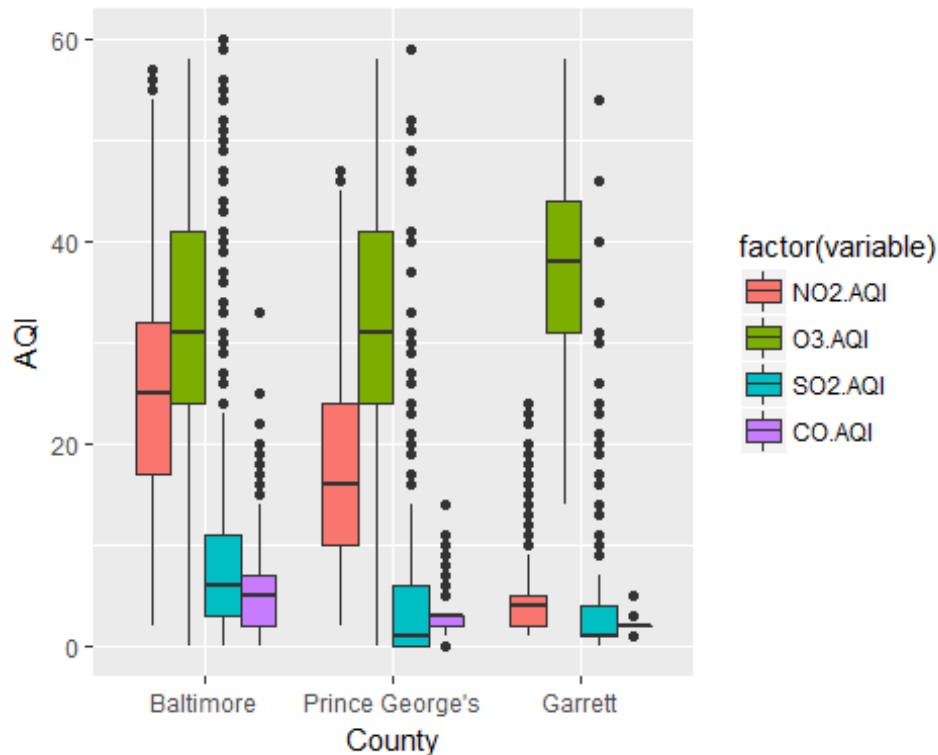
```
library(ggplot2)
library(reshape2)

# Convert County variable to factor.
pollution_md$County <- factor(pollution_md$County, levels = c("Baltimore",
"Prince George's", "Garrett"))

# Subset and transform the data for use in boxplot.
pollution_md2 <- pollution_md[, c("State", "County", "NO2.AQI", "O3.AQI",
"SO2.AQI", "CO.AQI")]
pollution_md3 <- melt(pollution_md2, id.vars = c("State", "County"),
measure.vars = c("NO2.AQI", "O3.AQI", "SO2.AQI", "CO.AQI"), na.rm = TRUE)

# Boxplot to show how AQI for each of the four pollutants is distributed
among counties in Maryland.
ggplot(data = pollution_md3) + geom_boxplot(aes(x = County, y = value, fill =
factor(variable))) + ylab("AQI") + ylim(0, 60)
```

```
## Warning: Removed 2130 rows containing non-finite values (stat_boxplot).
```



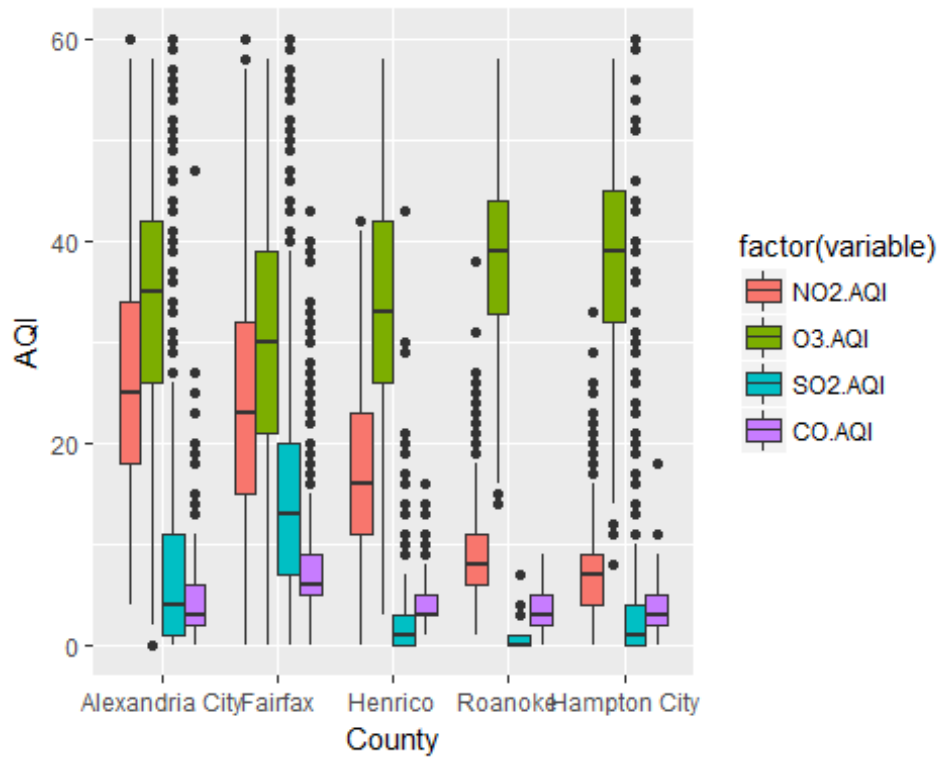
IV(b). Boxplot for counties in Virginia

```
# Convert County variable to factor
pollution_va$County <- factor(pollution_va$County,
                               levels = c("Alexandria City", "Fairfax",
                                           "Henrico", "Roanoke", "Hampton City"))

# Subset and transform the data for use in boxplot.
pollution_va2 <- pollution_va[, c("State", "County", "NO2.AQI", "O3.AQI",
                                   "SO2.AQI", "CO.AQI")]
pollution_va3 <- melt(pollution_va2, id.vars = c("State", "County"),
                      measure.vars = c("NO2.AQI", "O3.AQI", "SO2.AQI", "CO.AQI"), na.rm = TRUE)

# Boxplot to show how AQI for each of the four pollutants is distributed
# among counties in Virginia.
ggplot(data = pollution_va3) + geom_boxplot(aes(x = County, y = value, fill =
factor(variable))) + ylab("AQI") + ylim(0, 60)

## Warning: Removed 3872 rows containing non-finite values (stat_boxplot).
```

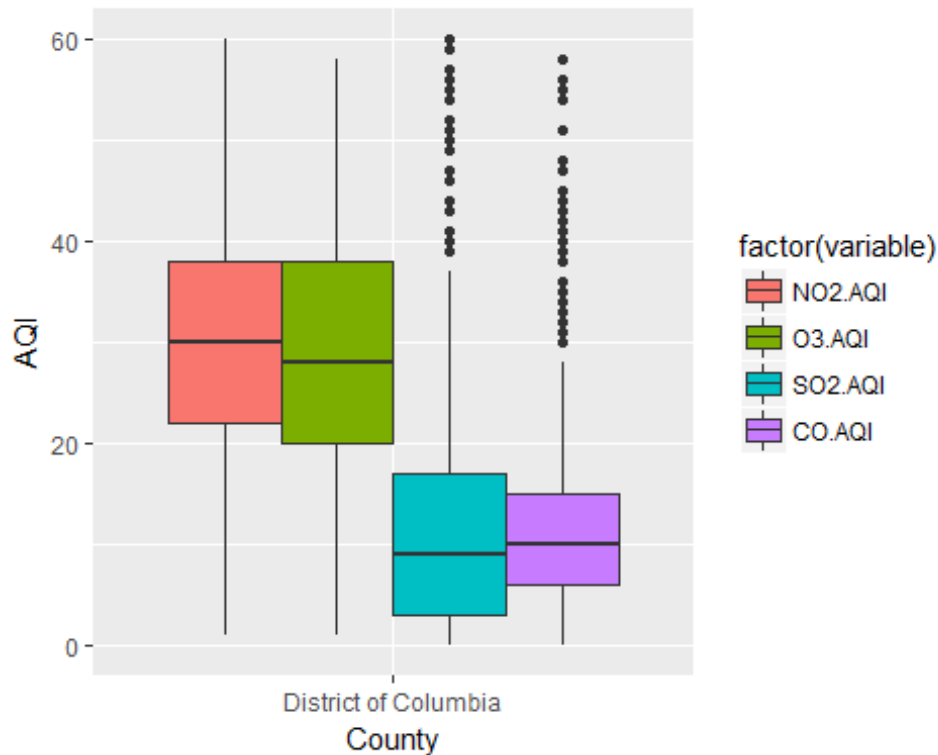


#### IV(c). Boxplot for Washington DC

```
# Subset and transform the data for use in boxplot.
pollution_dc2 <- pollution_dc[, c("State", "County", "NO2.AQI", "O3.AQI",
"SO2.AQI", "CO.AQI")]
pollution_dc3 <- melt(pollution_dc2, id.vars =c("State", "County"),
measure.vars = c("NO2.AQI", "O3.AQI", "SO2.AQI", "CO.AQI"), na.rm = TRUE)

# Boxplot to show how AQI for each of the four pollutants is distributed in
Washington DC.
ggplot(data = pollution_dc3) + geom_boxplot(aes(x = County, y = value, fill =
factor(variable))) + ylab("AQI") + ylim(0, 60)

## Warning: Removed 2682 rows containing non-finite values (stat_boxplot).
```



## V. ANOVA

We decided to run one-way ANOVA tests for counties in MD and VA to test for potential differences in pollutant AQIs. For each state, we have 4 separate tests with the pollutant AQI as our dependent variable and County as our independent variable. We then ran a one-way ANOVA test with State (MD, VA, DC) as the independent variable. Finally, to take a closer look at how pollutant AQIs compare at the county level for all states combined, we ran an ANOVA test on pollutant AQIs with County as the independent variable for a subset that includes MD, VA, and DC.

V(a). ANOVA test for counties in Maryland

```
par(mfrow = c(2, 2))

# One-way ANOVA, summary, and post-hoc test for NO2 AQI and counties in MD
plot(NO2.AQI ~ County, data = pollution_md)
aov_md_NO2 <- aov(NO2.AQI ~ County, data = pollution_md)
summary(aov_md_NO2)

##              Df  Sum Sq Mean Sq F value Pr(>F)
## County         2 1066884   533442    6108 <2e-16 ***
## Residuals    23535 2055328         87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_md_NO2 <- TukeyHSD(aov_md_NO2)
tukey_md_NO2
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = NO2.AQI ~ County, data = pollution_md)
##
## $County
##               diff            lwr            upr p adj
## Prince George's-Baltimore -7.333339 -7.651758 -7.014921 0
## Garrett-Baltimore -20.301506 -20.745816 -19.857197 0
## Garrett-Prince George's -12.968167 -13.443527 -12.492807 0

# One-way ANOVA, summary, and post-hoc test for O3 AQI and counties in MD
plot(O3.AQI ~ County, data = pollution_md)
aov_md_O3 <- aov(O3.AQI ~ County, data = pollution_md)
summary(aov_md_O3)

##              Df Sum Sq Mean Sq F value Pr(>F)
## County         2   17864    8932   24.86 1.65e-11 ***
## Residuals    23535  8456883     359
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_md_O3 <- TukeyHSD(aov_md_O3)
tukey_md_O3

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = O3.AQI ~ County, data = pollution_md)
##
## $County
##               diff            lwr            upr p adj
## Prince George's-Baltimore -0.1098125 -0.7557091 0.5360841 0.9162019
## Garrett-Baltimore         2.5761826  1.6749222 3.4774430 0.0000000
## Garrett-Prince George's    2.6859951  1.7217511 3.6502391 0.0000000

# One-way ANOVA, summary, and post-hoc test for SO2 AQI and counties in MD
plot(SO2.AQI ~ County, data = pollution_md)
aov_md_SO2 <- aov(SO2.AQI ~ County, data = pollution_md)
summary(aov_md_SO2)

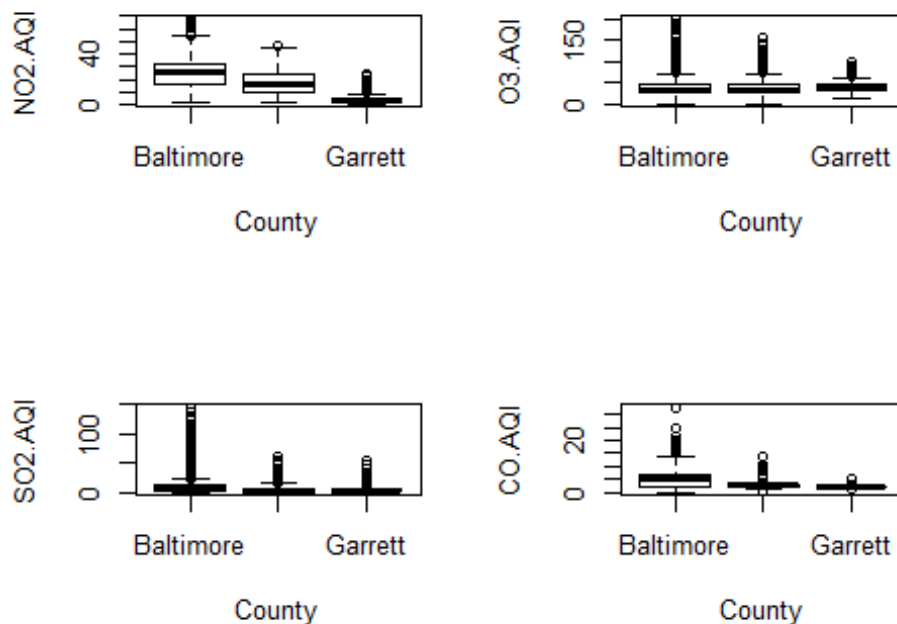
##              Df Sum Sq Mean Sq F value Pr(>F)
## County         2  103354    51677   440.7 <2e-16 ***
## Residuals    11769 1379942     117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 11766 observations deleted due to missingness

tukey_md_SO2 <- TukeyHSD(aov_md_SO2)
tukey_md_SO2
```



```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = S02.AQI ~ County, data = pollution_md)
##
## $County
##               diff      lwr      upr    p adj
## Prince George's-Baltimore -5.8323227 -6.354050 -5.3105957 0.0000000
## Garrett-Baltimore         -6.2842909 -7.012217 -5.5563651 0.0000000
## Garrett-Prince George's   -0.4519683 -1.230731  0.3267942 0.3619237

# One-way ANOVA, summary, and post-hoc test for CO AQI and counties in MD
plot(CO.AQI ~ County, data = pollution_md)
```



```
aov_md_CO <- aov(CO.AQI ~ County, data = pollution_md)
summary(aov_md_CO)

##               Df Sum Sq Mean Sq F value Pr(>F)
## County         2  22280   11140    1257 <2e-16 ***
## Residuals    11770 104279         9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 11765 observations deleted due to missingness

tukey_md_CO <- TukeyHSD(aov_md_CO)
tukey_md_CO
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = CO.AQI ~ County, data = pollution_md)
##
## $County
##
```

	diff	lwr	upr	p adj
Prince George's-Baltimore	-2.303917	-2.447337	-2.160497	0
Garrett-Baltimore	-3.560544	-3.760735	-3.360353	0
Garrett-Prince George's	-1.256627	-1.470816	-1.042439	0

V(b). ANOVA test for counties in Virginia

```
par(mfrow = c(2, 2))

# One-way ANOVA, summary, and post-hoc test for NO2 AQI and counties in VA
plot(NO2.AQI ~ County, data = pollution_va)
aov_va_NO2 <- aov(NO2.AQI ~ County, data = pollution_va)
summary(aov_va_NO2)

##
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
County	4	1569750	392438	3758	<2e-16 ***
Residuals	36417	3802484	104		

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_va_NO2 <- TukeyHSD(aov_va_NO2)
tukey_va_NO2

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = NO2.AQI ~ County, data = pollution_va)
##
## $County
##
```

	diff	lwr	upr	p adj
Fairfax-Alexandria City	-2.555304	-2.953744	-2.156864	0
Henrico-Alexandria City	-9.450774	-9.942731	-8.958817	0
Roanoke-Alexandria City	-17.089502	-17.775823	-16.403180	0
Hampton City-Alexandria City	-18.905337	-19.431148	-18.379525	0
Henrico-Fairfax	-6.895470	-7.315885	-6.475055	0
Roanoke-Fairfax	-14.534197	-15.171194	-13.897201	0
Hampton City-Fairfax	-16.350032	-16.809602	-15.890462	0
Roanoke-Henrico	-7.638727	-8.338035	-6.939420	0
Hampton City-Henrico	-9.454562	-9.997215	-8.911910	0
Hampton City-Roanoke	-1.815835	-2.539359	-1.092311	0

```
##
# One-way ANOVA, summary, and post-hoc test for SO2 AQI and counties in VA
plot(SO2.AQI ~ County, data = pollution_va)
aov_va_SO2 <- aov(SO2.AQI ~ County, data = pollution_va)
summary(aov_va_SO2)
```

```
##           Df  Sum Sq Mean Sq F value Pr(>F)
## County      4  617648  154412    1446 <2e-16 ***
## Residuals 18206 1943920    107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 18211 observations deleted due to missingness

tukey_va_S02 <- TukeyHSD(aov_va_S02)
tukey_va_S02

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = S02.AQI ~ County, data = pollution_va)
##
## $County
##           diff          lwr          upr      p adj
## Fairfax-Alexandria City    6.587101    6.017236    7.156966 0.0000000
## Henrico-Alexandria City   -6.964766   -7.668382   -6.261149 0.0000000
## Roanoke-Alexandria City   -8.445944   -9.427548   -7.464339 0.0000000
## Hampton City-Alexandria City -4.423338   -5.175375   -3.671302 0.0000000
## Henrico-Fairfax          -13.551866  -14.153161  -12.950572 0.0000000
## Roanoke-Fairfax          -15.033044  -15.944102  -14.121987 0.0000000
## Hampton City-Fairfax     -11.010439  -11.667735  -10.353144 0.0000000
## Roanoke-Henrico          -1.481178   -2.481356   -0.481000 0.0005146
## Hampton City-Henrico      2.541427    1.765304    3.317551 0.0000000
## Hampton City-Roanoke      4.022605    2.987792    5.057419 0.0000000

# One-way ANOVA, summary, and post-hoc test for O3 AQI and counties in VA
plot(O3.AQI ~ County, data = pollution_va)
aov_va_O3 <- aov(O3.AQI ~ County, data = pollution_va)
summary(aov_va_O3)

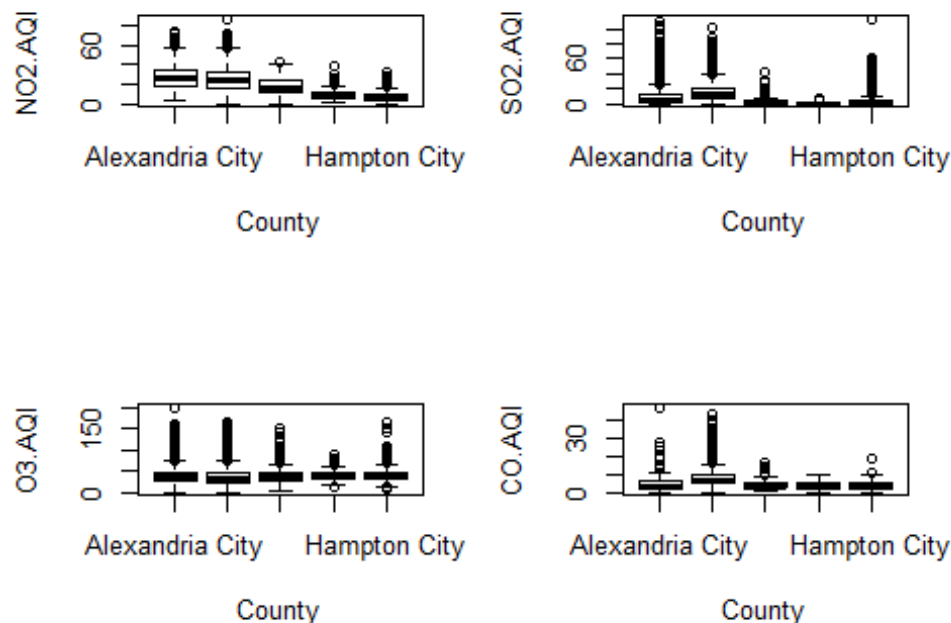
##           Df  Sum Sq Mean Sq F value Pr(>F)
## County      4  195624  48906    125 <2e-16 ***
## Residuals 36417 14243630    391
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_va_O3 <- TukeyHSD(aov_va_O3)
tukey_va_O3

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = O3.AQI ~ County, data = pollution_va)
##
## $County
##           diff          lwr          upr      p adj
## Fairfax-Alexandria City   -4.0237827  -4.7949343  -3.2526311 0.0000000
## Henrico-Alexandria City   -3.7638929  -4.7160391  -2.8117467 0.0000000
```

```
## Roanoke-Alexandria City      -0.5253194 -1.8536441  0.8030052 0.8175685
## Hampton City-Alexandria City 2.0689966  1.0513277  3.0866655 0.0000003
## Henrico-Fairfax             0.2598898 -0.5537924  1.0735719 0.9074885
## Roanoke-Fairfax             3.4984633  2.2656037  4.7313228 0.0000000
## Hampton City-Fairfax        6.0927793  5.2033153  6.9822433 0.0000000
## Roanoke-Henrico             3.2385735  1.8851149  4.5920322 0.0000000
## Hampton City-Henrico        5.8328896  4.7826258  6.8831533 0.0000000
## Hampton City-Roanoke        2.5943160  1.1939882  3.9946438 0.0000043
```

```
# One-way ANOVA, summary, and post-hoc test for CO AQI and counties in VA
plot(CO.AQI ~ County, data = pollution_va)
```



```
aov_va_CO <- aov(CO.AQI ~ County, data = pollution_va)
summary(aov_va_CO)

##              Df Sum Sq Mean Sq F value Pr(>F)
## County         4  54079   13520    1190 <2e-16 ***
## Residuals    18209 206828      11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 18208 observations deleted due to missingness

tukey_va_CO <- TukeyHSD(aov_va_CO)
tukey_va_CO

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
```

```
## Fit: aov(formula = CO.AQI ~ County, data = pollution_va)
##
## $County
##              diff          lwr          upr          p adj
## Fairfax-Alexandria City    3.1739841  2.9881272  3.35984107 0.0000000
## Henrico-Alexandria City   -0.2033810 -0.4328720  0.02611005 0.1105815
## Roanoke-Alexandria City   -0.7553411 -1.0755006 -0.43518168 0.0000000
## Hampton City-Alexandria City -0.5052523 -0.7505360 -0.25996867 0.0000002
## Henrico-Fairfax          -3.3773651 -3.5734735 -3.18125674 0.0000000
## Roanoke-Fairfax          -3.9293253 -4.2264691 -3.63218146 0.0000000
## Hampton City-Fairfax     -3.6792365 -3.8936110 -3.46486194 0.0000000
## Roanoke-Henrico          -0.5519602 -0.8781775 -0.22574276 0.0000388
## Hampton City-Henrico     -0.3018713 -0.5550112 -0.04873150 0.0100656
## Hampton City-Roanoke      0.2500888 -0.0874252  0.58760285 0.2556324
```

V(c). ANOVA test for the states of Maryland, Virginia, and Washington DC

```
par(mfrow = c(2, 2))

# One-way ANOVA, summary, and post-hoc test for NO2 AQI and the states of MD,
# VA, DC
plot(NO2.AQI ~ State, data = pollution_tri)
aov_tri_NO2 <- aov(NO2.AQI ~ State, data = pollution_tri)
summary(aov_tri_NO2)

##              Df    Sum Sq Mean Sq F value Pr(>F)
## State          2  1997686   998843    6691 <2e-16 ***
## Residuals    85653 12786566     149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_tri_NO2 <- TukeyHSD(aov_tri_NO2)
tukey_tri_NO2

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = NO2.AQI ~ State, data = pollution_tri)
##
## $State
##              diff          lwr          upr p adj
## Maryland-District Of Columbia -10.8244945 -11.0828531 -10.5661360 0e+00
## Virginia-District Of Columbia -10.3390439 -10.5723370 -10.1057508 0e+00
## Virginia-Maryland              0.4854506   0.2459691   0.7249321 6e-06

# One-way ANOVA, summary, and post-hoc test for SO2 AQI and the states of MD,
# VA, DC
plot(SO2.AQI ~ State, data = pollution_tri)
aov_tri_SO2 <- aov(SO2.AQI ~ State, data = pollution_tri)
summary(aov_tri_SO2)
```

```
##              Df  Sum Sq Mean Sq F value Pr(>F)
## State          2   209183   104591    712.7 <2e-16 ***
## Residuals    42841 6286856     147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 42812 observations deleted due to missingness

tukey_tri_S02 <- TukeyHSD(aov_tri_S02)
tukey_tri_S02

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = S02.AQI ~ State, data = pollution_tri)
##
## $State
##              diff            lwr            upr p adj
## Maryland-District Of Columbia -5.82467 -6.186817 -5.462523    0
## Virginia-District Of Columbia -3.03490 -3.361916 -2.707884    0
## Virginia-Maryland              2.78977  2.454006  3.125534    0

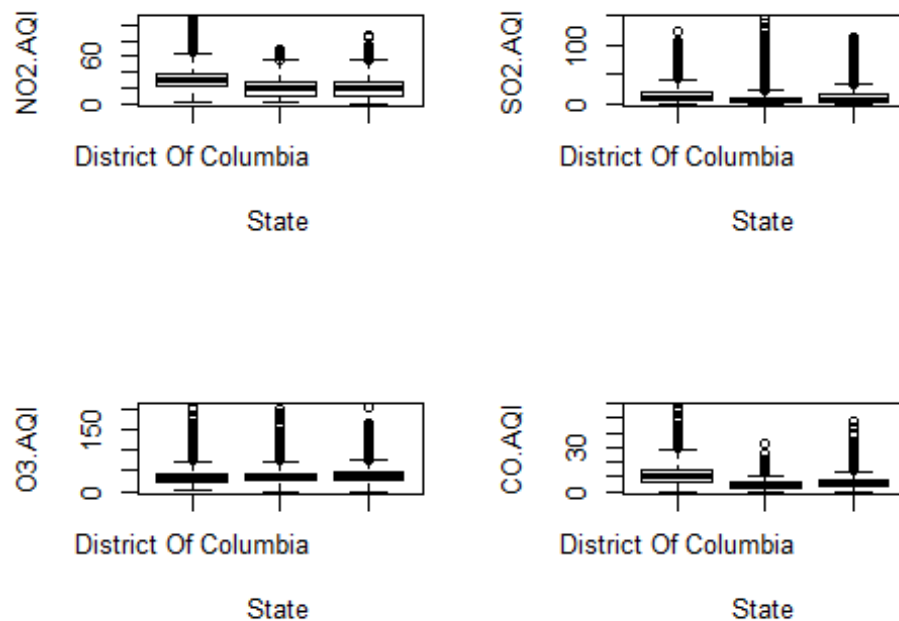
# One-way ANOVA, summary, and post-hoc test for O3 AQI and the states of MD,
# VA, DC
plot(O3.AQI ~ State, data = pollution_tri)
aov_tri_O3 <- aov(O3.AQI ~ State, data = pollution_tri)
summary(aov_tri_O3)

##              Df  Sum Sq Mean Sq F value Pr(>F)
## State          2   281062   140531    352.1 <2e-16 ***
## Residuals    85653 34189304     399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_tri_O3 <- TukeyHSD(aov_tri_O3)
tukey_tri_O3

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = O3.AQI ~ State, data = pollution_tri)
##
## $State
##              diff            lwr            upr p adj
## Maryland-District Of Columbia 3.079872 2.6574065 3.502337    0
## Virginia-District Of Columbia 4.275874 3.8943959 4.657353    0
## Virginia-Maryland            1.196003 0.8044049 1.587600    0

# One-way ANOVA, summary, and post-hoc test for CO AQI and the states of MD,
# VA, DC
plot(CO.AQI ~ State, data = pollution_tri)
```



```
aov_tri_CO <- aov(CO.AQI ~ State, data = pollution_tri)
summary(aov_tri_CO)

##              Df Sum Sq Mean Sq F value Pr(>F)
## State          2  402839    201420     7636 <2e-16 ***
## Residuals    42833 1129867         26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 42820 observations deleted due to missingness

tukey_tri_CO <- TukeyHSD(aov_tri_CO)
tukey_tri_CO

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = CO.AQI ~ State, data = pollution_tri)
##
## $State
##              diff              lwr              upr p adj
## Maryland-District Of Columbia -7.2348515 -7.3884228 -7.081280      0
## Virginia-District Of Columbia -6.2452724 -6.3839514 -6.106593      0
## Virginia-Maryland              0.9895791  0.8472326  1.131926      0
```

V(d). ANOVA test for counties in Maryland, Virginia, and Washington DC

```
par(mfrow = c(2, 2))
```

*# One-way ANOVA, summary, and post-hoc test for NO2 AQI and counties in MD, VA, DC*

```
plot(NO2.AQI ~ County, data = pollution_tri)
```

```
aov_tri_NO2 <- aov(NO2.AQI ~ County, data = pollution_tri)
```

```
summary(aov_tri_NO2)
```

```
##              Df    Sum Sq Mean Sq F value Pr(>F)
## County          8  4634321   579290    4888 <2e-16 ***
## Residuals    85647 10149932     119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tukey_tri_NO2 <- TukeyHSD(aov_tri_NO2)
```

```
tukey_tri_NO2
```

```
## Tukey multiple comparisons of means
```

```
## 95% family-wise confidence level
```

```
##
```

```
## Fit: aov(formula = NO2.AQI ~ County, data = pollution_tri)
```

```
##
```

```
## $County
```

```
##              diff              lwr              upr
## Baltimore-Alexandria City    -1.7989395    -2.2998639    -1.2980150
## District of Columbia-Alexandria City  4.1547985     3.6979059     4.6116911
## Fairfax-Alexandria City      -2.5553042    -3.0379767    -2.0726317
## Garrett-Alexandria City      -22.1004459   -22.8400258   -21.3608659
## Hampton City-Alexandria City -18.9053365   -19.5423070   -18.2683660
## Henrico-Alexandria City      -9.4507743   -10.0467334    -8.8548151
## Prince George's-Alexandria City -9.1322789    -9.6968916    -8.5676662
## Roanoke-Alexandria City      -17.0895017   -17.9209152   -16.2580882
## District of Columbia-Baltimore  5.9537380     5.5919158     6.3155602
## Fairfax-Baltimore           -0.7563647    -1.1502394    -0.3624901
## Garrett-Baltimore           -20.3015064   -20.9864466   -19.6165663
## Hampton City-Baltimore       -17.1063971   -17.6790190   -16.5337751
## Henrico-Baltimore           -7.6518348    -8.1784578    -7.1252118
## Prince George's-Baltimore     -7.3333394    -7.8242081    -6.8424708
## Roanoke-Baltimore           -15.2905622   -16.0737690   -14.5073554
## Fairfax-District of Columbia  -6.7101027    -7.0462021    -6.3740033
## Garrett-District of Columbia -26.2552444   -26.9086727   -25.6018161
## Hampton City-District of Columbia -23.0601350   -23.5946642   -22.5256058
## Henrico-District of Columbia -13.6055728   -14.0905030   -13.1206426
## Prince George's-District of Columbia -13.2870774   -13.7329222   -12.8412326
## Roanoke-District of Columbia -21.2443002   -22.0001034   -20.4884970
## Garrett-Fairfax             -19.5451417   -20.2168488   -18.8734345
## Hampton City-Fairfax         -16.3500323   -16.9067579   -15.7933067
## Henrico-Fairfax             -6.8954701    -7.4047629    -6.3861772
## Prince George's-Fairfax       -6.5769747    -7.0492029    -6.1047465
## Roanoke-Fairfax             -14.5341975   -15.3058583   -13.7625367
## Hampton City-Garrett         3.1951094     2.4052048     3.9850139
## Henrico-Garrett             12.6496716    11.8924497    13.4068935
```



## Prince George's-Garrett	12.9681670	12.2353606	13.7009734
## Roanoke-Garrett	5.0109442	4.0572849	5.9646035
## Henrico-Hampton City	9.4545623	8.7971903	10.1119342
## Prince George's-Hampton City	9.7730576	9.1439645	10.4021508
## Roanoke-Hampton City	1.8158348	0.9393538	2.6923159
## Prince George's-Henrico	0.3184954	-0.2690368	0.9060275
## Roanoke-Henrico	-7.6387274	-8.4858726	-6.7915823
## Roanoke-Prince George's	-7.9572228	-8.7826167	-7.1318288
##	p adj		
## Baltimore-Alexandria City	0.0000000		
## District of Columbia-Alexandria City	0.0000000		
## Fairfax-Alexandria City	0.0000000		
## Garrett-Alexandria City	0.0000000		
## Hampton City-Alexandria City	0.0000000		
## Henrico-Alexandria City	0.0000000		
## Prince George's-Alexandria City	0.0000000		
## Roanoke-Alexandria City	0.0000000		
## District of Columbia-Baltimore	0.0000000		
## Fairfax-Baltimore	0.0000001		
## Garrett-Baltimore	0.0000000		
## Hampton City-Baltimore	0.0000000		
## Henrico-Baltimore	0.0000000		
## Prince George's-Baltimore	0.0000000		
## Roanoke-Baltimore	0.0000000		
## Fairfax-District of Columbia	0.0000000		
## Garrett-District of Columbia	0.0000000		
## Hampton City-District of Columbia	0.0000000		
## Henrico-District of Columbia	0.0000000		
## Prince George's-District of Columbia	0.0000000		
## Roanoke-District of Columbia	0.0000000		
## Garrett-Fairfax	0.0000000		
## Hampton City-Fairfax	0.0000000		
## Henrico-Fairfax	0.0000000		
## Prince George's-Fairfax	0.0000000		
## Roanoke-Fairfax	0.0000000		
## Hampton City-Garrett	0.0000000		
## Henrico-Garrett	0.0000000		
## Prince George's-Garrett	0.0000000		
## Roanoke-Garrett	0.0000000		
## Henrico-Hampton City	0.0000000		
## Prince George's-Hampton City	0.0000000		
## Roanoke-Hampton City	0.0000000		
## Prince George's-Henrico	0.7581780		
## Roanoke-Henrico	0.0000000		
## Roanoke-Prince George's	0.0000000		

*# One-way ANOVA, summary, and post-hoc test for SO2 AQI and counties in MD, VA, DC*

```
plot(SO2.AQI ~ County, data = pollution_tri)
```

```

aov_tri_S02 <- aov(S02.AQI ~ County, data = pollution_tri)
summary(aov_tri_S02)

##              Df Sum Sq Mean Sq F value Pr(>F)
## County          8  930185   116273   894.8 <2e-16 ***
## Residuals    42835 5565854     130
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 42812 observations deleted due to missingness

tukey_tri_S02 <- TukeyHSD(aov_tri_S02)
tukey_tri_S02

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = S02.AQI ~ County, data = pollution_tri)
##
## $County
##
##              diff              lwr              upr
## Baltimore-Alexandria City    0.6242209 -0.1175644  1.3660063
## District of Columbia-Alexandria City  3.8227886  3.1462798  4.4992974
## Fairfax-Alexandria City    6.5871008  5.8723435  7.3018581
## Garrett-Alexandria City   -5.6600700 -6.7550073 -4.5651326
## Hampton City-Alexandria City -4.4233383 -5.3665851 -3.4800914
## Henrico-Alexandria City   -6.9647656 -7.8472815 -6.0822498
## Prince George's-Alexandria City -5.2081017 -6.0440892 -4.3721143
## Roanoke-Alexandria City   -8.4459436 -9.6771280 -7.2147592
## District of Columbia-Baltimore  3.1985676  2.6628612  3.7342741
## Fairfax-Baltimore        5.9628799  5.3796173  6.5461424
## Garrett-Baltimore       -6.2842909 -7.2982954 -5.2702865
## Hampton City-Baltimore   -5.0475592 -5.8955167 -4.1996018
## Henrico-Baltimore       -7.5889866 -8.3688272 -6.8091460
## Prince George's-Baltimore -5.8323227 -6.5590910 -5.1055544
## Roanoke-Baltimore       -9.0701645 -10.2299629 -7.9103662
## Fairfax-District of Columbia  2.7643122  2.2667041  3.2619204
## Garrett-District of Columbia -9.4828586 -10.4501350 -8.5155822
## Hampton City-District of Columbia -8.2461269 -9.0376132 -7.4546406
## Henrico-District of Columbia -10.7875542 -11.5055863 -10.0695221
## Prince George's-District of Columbia -9.0308903 -9.6908986 -8.3708821
## Roanoke-District of Columbia -12.2687322 -13.3879064 -11.1495580
## Garrett-Fairfax         -12.2471708 -13.2415739 -11.2527677
## Hampton City-Fairfax    -11.0104391 -11.8348567 -10.1860215
## Henrico-Fairfax        -13.5518665 -14.3060441 -12.7976888
## Prince George's-Fairfax -11.7952026 -12.4943626 -11.0960426
## Roanoke-Fairfax        -15.0330444 -16.1757451 -13.8903438
## Hampton City-Garrett     1.2367317  0.0672557  2.4062077
## Henrico-Garrett        -1.3046956 -2.4257637 -0.1836275
## Prince George's-Garrett  0.4519683 -0.6328518  1.5367883
## Roanoke-Garrett        -2.7858736 -4.1978844 -1.3738628

```

## Henrico-Hampton City	-2.5414274	-3.5148854	-1.5679693
## Prince George's-Hampton City	-0.7847634	-1.7162469	0.1467200
## Roanoke-Hampton City	-4.0226053	-5.3205272	-2.7246834
## Prince George's-Henrico	1.7566639	0.8867324	2.6265954
## Roanoke-Henrico	-1.4811779	-2.7356583	-0.2266976
## Roanoke-Prince George's	-3.2378419	-4.4600373	-2.0156464
##	p adj		
## Baltimore-Alexandria City	0.1821340		
## District of Columbia-Alexandria City	0.0000000		
## Fairfax-Alexandria City	0.0000000		
## Garrett-Alexandria City	0.0000000		
## Hampton City-Alexandria City	0.0000000		
## Henrico-Alexandria City	0.0000000		
## Prince George's-Alexandria City	0.0000000		
## Roanoke-Alexandria City	0.0000000		
## District of Columbia-Baltimore	0.0000000		
## Fairfax-Baltimore	0.0000000		
## Garrett-Baltimore	0.0000000		
## Hampton City-Baltimore	0.0000000		
## Henrico-Baltimore	0.0000000		
## Prince George's-Baltimore	0.0000000		
## Roanoke-Baltimore	0.0000000		
## Fairfax-District of Columbia	0.0000000		
## Garrett-District of Columbia	0.0000000		
## Hampton City-District of Columbia	0.0000000		
## Henrico-District of Columbia	0.0000000		
## Prince George's-District of Columbia	0.0000000		
## Roanoke-District of Columbia	0.0000000		
## Garrett-Fairfax	0.0000000		
## Hampton City-Fairfax	0.0000000		
## Henrico-Fairfax	0.0000000		
## Prince George's-Fairfax	0.0000000		
## Roanoke-Fairfax	0.0000000		
## Hampton City-Garrett	0.0287915		
## Henrico-Garrett	0.0093214		
## Prince George's-Garrett	0.9339155		
## Roanoke-Garrett	0.0000000		
## Henrico-Hampton City	0.0000000		
## Prince George's-Hampton City	0.1808847		
## Roanoke-Hampton City	0.0000000		
## Prince George's-Henrico	0.0000000		
## Roanoke-Henrico	0.0076942		
## Roanoke-Prince George's	0.0000000		

*# One-way ANOVA, summary, and post-hoc test for O3 AQI and counties in MD, VA, DC*

```
plot(O3.AQI ~ County, data = pollution_tri)
aov_tri_O3 <- aov(O3.AQI ~ County, data = pollution_tri)
summary(aov_tri_O3)
```

```
##           Df    Sum Sq Mean Sq F value Pr(>F)
## County      8    494550    61819   155.8 <2e-16 ***
## Residuals 85647 33975816     397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_tri_03 <- TukeyHSD(aov_tri_03)
tukey_tri_03

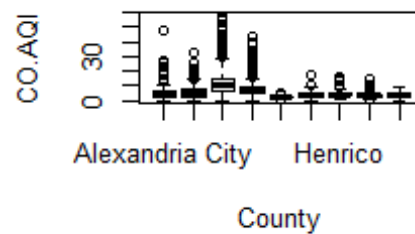
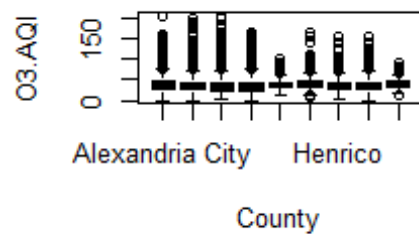
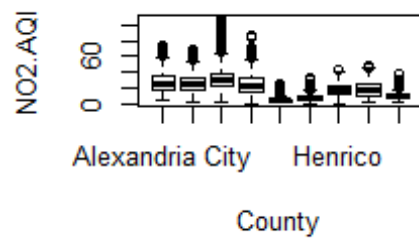
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = O3.AQI ~ County, data = pollution_tri)
##
## $County
##
```

	diff	lwr	upr
Baltimore-Alexandria City	-3.70447812	-4.6209636	-2.7879926
District of Columbia-Alexandria City	-6.49264419	-7.3285695	-5.6567189
Fairfax-Alexandria City	-4.02378270	-4.9068747	-3.1406907
Garrett-Alexandria City	-1.12829552	-2.4814224	0.2248313
Hampton City-Alexandria City	2.06899662	0.9036028	3.2343904
Henrico-Alexandria City	-3.76389294	-4.8542528	-2.6735331
Prince George's-Alexandria City	-3.81429063	-4.8472995	-2.7812818
Roanoke-Alexandria City	-0.52531942	-2.0464639	0.9958250
District of Columbia-Baltimore	-2.78816607	-3.4501518	-2.1261804
Fairfax-Baltimore	-0.31930458	-1.0399331	0.4013239
Garrett-Baltimore	2.57618260	1.3230241	3.8293411
Hampton City-Baltimore	5.77347473	4.7258122	6.8211372
Henrico-Baltimore	-0.05941483	-1.0229181	0.9040884
Prince George's-Baltimore	-0.10981252	-1.0079001	0.7882750
Roanoke-Baltimore	3.17915869	1.7462127	4.6121047
Fairfax-District of Columbia	2.46886149	1.8539380	3.0837850
Garrett-District of Columbia	5.36434867	4.1688439	6.5598534
Hampton City-District of Columbia	8.56164080	7.5836724	9.5396092
Henrico-District of Columbia	2.72875124	1.8415286	3.6159739
Prince George's-District of Columbia	2.67835355	1.8626411	3.4940660
Roanoke-District of Columbia	5.96732476	4.5845161	7.3501334
Garrett-Fairfax	2.89548718	1.6665396	4.1244347
Hampton City-Fairfax	6.09277932	5.0742006	7.1113580
Henrico-Fairfax	0.25988976	-0.6719066	1.1916861
Prince George's-Fairfax	0.20949207	-0.6544912	1.0734754
Roanoke-Fairfax	3.49846328	2.0866417	4.9102849
Hampton City-Garrett	3.19729214	1.7520920	4.6424922
Henrico-Garrett	-2.63559742	-4.0210018	-1.2501931
Prince George's-Garrett	-2.68599511	-4.0267292	-1.3452610
Roanoke-Garrett	0.60297609	-1.1418279	2.3477801
Henrico-Hampton City	-5.83288956	-7.0356097	-4.6301694
Prince George's-Hampton City	-5.88328725	-7.0342687	-4.7323058
Roanoke-Hampton City	-2.59431604	-4.1979156	-0.9907165
Prince George's-Henrico	-0.05039769	-1.1253397	1.0245443

```
## Roanoke-Henrico 3.23857352 1.6886467 4.7885004
## Roanoke-Prince George's 3.28897121 1.7788401 4.7991023
## p adj
## Baltimore-Alexandria City 0.0000000
## District of Columbia-Alexandria City 0.0000000
## Fairfax-Alexandria City 0.0000000
## Garrett-Alexandria City 0.1921589
## Hampton City-Alexandria City 0.0000013
## Henrico-Alexandria City 0.0000000
## Prince George's-Alexandria City 0.0000000
## Roanoke-Alexandria City 0.9783046
## District of Columbia-Baltimore 0.0000000
## Fairfax-Baltimore 0.9076618
## Garrett-Baltimore 0.0000000
## Hampton City-Baltimore 0.0000000
## Henrico-Baltimore 0.9999999
## Prince George's-Baltimore 0.9999883
## Roanoke-Baltimore 0.0000000
## Fairfax-District of Columbia 0.0000000
## Garrett-District of Columbia 0.0000000
## Hampton City-District of Columbia 0.0000000
## Henrico-District of Columbia 0.0000000
## Prince George's-District of Columbia 0.0000000
## Roanoke-District of Columbia 0.0000000
## Garrett-Fairfax 0.0000000
## Hampton City-Fairfax 0.0000000
## Henrico-Fairfax 0.9946708
## Prince George's-Fairfax 0.9979948
## Roanoke-Fairfax 0.0000000
## Hampton City-Garrett 0.0000000
## Henrico-Garrett 0.0000001
## Prince George's-Garrett 0.0000000
## Roanoke-Garrett 0.9782103
## Henrico-Hampton City 0.0000000
## Prince George's-Hampton City 0.0000000
## Roanoke-Hampton City 0.0000185
## Prince George's-Henrico 1.0000000
## Roanoke-Henrico 0.0000000
## Roanoke-Prince George's 0.0000000
```

*# One-way ANOVA, summary, and post-hoc test for CO AQI and counties in MD, VA, DC*

```
plot(CO.AQI ~ County, data = pollution_tri)
```



```
aov_tri_CO <- aov(CO.AQI ~ County, data = pollution_tri)
summary(aov_tri_CO)

##              Df Sum Sq Mean Sq F value Pr(>F)
## County         8  479199    59900   2435 <2e-16 ***
## Residuals    42827 1053508         25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 42820 observations deleted due to missingness

tukey_tri_CO <- TukeyHSD(aov_tri_CO)
tukey_tri_CO

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = CO.AQI ~ County, data = pollution_tri)
##
## $County
##              diff              lwr              upr
## Baltimore-Alexandria City    1.48829590    1.1655841    1.811007754
## District of Columbia-Alexandria City  7.55031884    7.2559377    7.844699996
## Fairfax-Alexandria City      3.17398413    2.8630066    3.484961666
## Garrett-Alexandria City     -2.07224800   -2.5488838   -1.595612233
## Hampton City-Alexandria City -0.50525233   -0.9156632   -0.094841460
## Henrico-Alexandria City      -0.20338100   -0.5873675    0.180605529
## Prince George's-Alexandria City -0.81562086   -1.1794104   -0.451831334
```

## Roanoke-Alexandria City	-0.75534115	-1.2910349	-0.219647386
## District of Columbia-Baltimore	6.06202293	5.8289563	6.295089565
## Fairfax-Baltimore	1.68568822	1.4319821	1.939394386
## Garrett-Baltimore	-3.56054391	-4.0019524	-3.119135429
## Hampton City-Baltimore	-1.99354823	-2.3624613	-1.624635141
## Henrico-Baltimore	-1.69167690	-2.0309488	-1.352404975
## Prince George's-Baltimore	-2.30391676	-2.6201486	-1.987684877
## Roanoke-Baltimore	-2.24363705	-2.7482434	-1.739030677
## Fairfax-District of Columbia	-4.37633471	-4.5928622	-4.159807234
## Garrett-District of Columbia	-9.62256684	-10.0437067	-9.201426986
## Hampton City-District of Columbia	-8.05557116	-8.3999753	-7.711167043
## Henrico-District of Columbia	-7.75369983	-8.0661463	-7.441253406
## Prince George's-District of Columbia	-8.36593969	-8.6532025	-8.078676861
## Roanoke-District of Columbia	-8.30565998	-8.7926353	-7.818684703
## Garrett-Fairfax	-5.24623213	-5.6791357	-4.813328520
## Hampton City-Fairfax	-3.67923646	-4.0379299	-3.320543050
## Henrico-Fairfax	-3.37736513	-3.7054955	-3.049234769
## Prince George's-Fairfax	-3.98960498	-4.2938527	-3.685357231
## Roanoke-Fairfax	-3.92932528	-4.4265090	-3.432141522
## Hampton City-Garrett	1.56699568	1.0579421	2.076049296
## Henrico-Garrett	1.86886701	1.3808668	2.356867170
## Prince George's-Garrett	1.25662715	0.7843546	1.728899658
## Roanoke-Garrett	1.31690685	0.7023612	1.931452471
## Henrico-Hampton City	0.30187133	-0.1216845	0.725427208
## Prince George's-Hampton City	-0.31036853	-0.7157039	0.094966822
## Roanoke-Hampton City	-0.25008882	-0.8148204	0.314642741
## Prince George's-Henrico	-0.61223986	-0.9907967	-0.233682969
## Roanoke-Henrico	-0.55196015	-1.0977901	-0.006130224
## Roanoke-Prince George's	0.06027971	-0.4715356	0.592094970
##	p adj		
## Baltimore-Alexandria City	0.0000000		
## District of Columbia-Alexandria City	0.0000000		
## Fairfax-Alexandria City	0.0000000		
## Garrett-Alexandria City	0.0000000		
## Hampton City-Alexandria City	0.0042611		
## Henrico-Alexandria City	0.7810455		
## Prince George's-Alexandria City	0.0000000		
## Roanoke-Alexandria City	0.0004167		
## District of Columbia-Baltimore	0.0000000		
## Fairfax-Baltimore	0.0000000		
## Garrett-Baltimore	0.0000000		
## Hampton City-Baltimore	0.0000000		
## Henrico-Baltimore	0.0000000		
## Prince George's-Baltimore	0.0000000		
## Roanoke-Baltimore	0.0000000		
## Fairfax-District of Columbia	0.0000000		
## Garrett-District of Columbia	0.0000000		
## Hampton City-District of Columbia	0.0000000		
## Henrico-District of Columbia	0.0000000		
## Prince George's-District of Columbia	0.0000000		

## Roanoke-District of Columbia	0.0000000
## Garrett-Fairfax	0.0000000
## Hampton City-Fairfax	0.0000000
## Henrico-Fairfax	0.0000000
## Prince George's-Fairfax	0.0000000
## Roanoke-Fairfax	0.0000000
## Hampton City-Garrett	0.0000000
## Henrico-Garrett	0.0000000
## Prince George's-Garrett	0.0000000
## Roanoke-Garrett	0.0000000
## Henrico-Hampton City	0.3985298
## Prince George's-Hampton City	0.2976927
## Roanoke-Hampton City	0.9079332
## Prince George's-Henrico	0.0000186
## Roanoke-Henrico	0.0450404
## Roanoke-Prince George's	0.9999935

## VI. Correlation between variables (pollutant AQIs)

To see if there was any correlation between the pollutant AQIs themselves, we developed a scatterplot matrix for NO2, O3, SO2, and CO, then ran Pearson's Correlation for each pair of pollutant AQIs to see which could have the highest correlation, visualized by a correlation matrix. Tukey HSD Test was run for each of these compare the pairs of means within the selected groups.

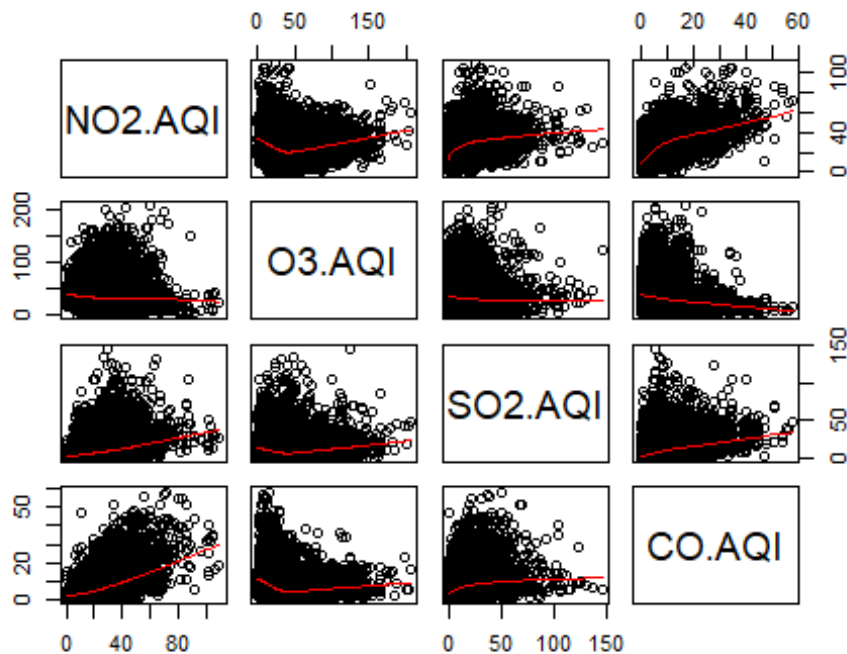
```
library(corrplot)

## Warning: package 'corrplot' was built under R version 3.4.3

## corrplot 0.84 loaded

# Scatterplot matrix for the pollutant AQIs (NO2, O3, SO2, CO)
pairs(~ NO2.AQI + O3.AQI + SO2.AQI + CO.AQI, data = pollution_tri,
panel=panel.smooth)
```





*# Pearson's Correlation for each pair of pollutant AQIs*

```
cor.test(pollution_tri$NO2.AQI, pollution_tri$CO.AQI)
```

```
##
## Pearson's product-moment correlation
##
## data: pollution_tri$NO2.AQI and pollution_tri$CO.AQI
## t = 153.49, df = 42834, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.5895508 0.6017702
## sample estimates:
## cor
## 0.5956949
```

```
cor.test(pollution_tri$NO2.AQI, pollution_tri$O3.AQI)
```

```
##
## Pearson's product-moment correlation
##
## data: pollution_tri$NO2.AQI and pollution_tri$O3.AQI
## t = -8.245, df = 85654, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.03485102 -0.02146795
## sample estimates:
```

```

##          cor
## -0.02816074

cor.test(pollution_tri$NO2.AQI, pollution_tri$SO2.AQI)

##
## Pearson's product-moment correlation
##
## data:  pollution_tri$NO2.AQI and pollution_tri$SO2.AQI
## t = 93.195, df = 42842, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4026528 0.4183990
## sample estimates:
##          cor
## 0.4105565

cor.test(pollution_tri$SO2.AQI, pollution_tri$O3.AQI)

##
## Pearson's product-moment correlation
##
## data:  pollution_tri$SO2.AQI and pollution_tri$O3.AQI
## t = -0.027839, df = 42842, p-value = 0.9778
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.009603513 0.009334539
## sample estimates:
##          cor
## -0.0001344991

cor.test(pollution_tri$SO2.AQI, pollution_tri$CO.AQI)

##
## Pearson's product-moment correlation
##
## data:  pollution_tri$SO2.AQI and pollution_tri$CO.AQI
## t = 66.074, df = 21424, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4002538 0.4225012
## sample estimates:
##          cor
## 0.4114388

cor.test(pollution_tri$O3.AQI, pollution_tri$CO.AQI)

##
## Pearson's product-moment correlation
##
## data:  pollution_tri$O3.AQI and pollution_tri$CO.AQI
## t = -36.232, df = 42834, p-value < 2.2e-16

```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.1816168 -0.1632401
## sample estimates:
##      cor
## -0.1724435

# Correlation matrix for pollutant AQIs
pollution_cor <- cor(pollution_tri[, c("NO2.AQI", "O3.AQI", "SO2.AQI",
"CO.AQI")],
                      method = "pearson", use = "complete.obs")
pollution_cor

##           NO2.AQI      O3.AQI      SO2.AQI      CO.AQI
## NO2.AQI  1.00000000 -0.0281918799  0.4105688434  0.5957083
## O3.AQI   -0.02819188  1.0000000000 -0.0002641859 -0.1725070
## SO2.AQI   0.41056884 -0.0002641859  1.0000000000  0.4114388
## CO.AQI    0.59570825 -0.1725070341  0.4114387848  1.0000000

# Correlation plot for pollutant AQIs
corrplot(pollution_cor, method = "circle")
```

