Trends in Alternative Daily Cover in California

https://github.com/sarahko7/ADC_Analysis

Sarah Ko

Abstract

Experimental overview. This section should be no longer than 250 words.

Contents

List of Tables

List of Figures

<Note: set up autoreferencing for figures and tables in your document>

1 Research Question and Rationale

2 Dataset Information

Column Name	Data Description
Report Year	Year that the ADC was used
Report Quarter	Quarter that the ADC was used
Ash	Ash and cement kiln dust materials
Auto Shredder Waste	Treated auto shredder waste
Construction and Demolition Waste	Processed construction and demolition wastes and materials
Compost	Compost materials
Contaminated Sediment	Contaminated sediment, dredge spoils, foundry sands
Green Material	Processed green material
Mixed	Mixtures of the other categories
Other	Before 1998, most ADC was classified in this category
Tires	Shredded tires
Sludge	Sludge and sludge-derived materials
Total	Sum of the columns Ash:Sludge

SKOTESTThis is the caption for the table Table 1: Summary of Data Structure

SKOTESTreference the table in text: Table 1

3 Exploratory Data Analysis and Wrangling

3.1 Data Wrangling

```
# per the CalRecycle website, segregation into ADC types started in 1998
# therefore, for the analysis, remove data from before 1998
class(ADC_raw$Report.Year)
```

```
## [1] "integer"
```

```
ADC_data <- filter(ADC_raw, Report.Year >= 1998)
dim(ADC_data)
```

```
## [1] 80 13
```

```
# explore new dataset
head(ADC_data)
```

```
##   Report.Year Report.Quarter      Ash Auto.Shredder.Waste
## 1        2017              1 32511.83            153270.6
## 2        2017              2 37294.78            159759.7
## 3        2017              3 33349.25            153342.6
## 4        2017              4 22248.85            123203.5
## 5        2016              1 31423.40            123193.3
## 6        2016              2 45504.45            126040.9
##   Construction.and.Demolition.Waste  Compost Contaminated.Sediment
## 1                          173548.6  6128.89               3396.36
## 2                          199486.8  2746.22               7585.58
## 3                          164028.4  1796.97               4280.92
## 4                          198901.7 15993.13               2979.12
## 5                          160446.5 15681.63              20203.18
## 6                          144982.9 42215.62              18089.73
##   Green.Material    Mixed    Other   Tires    Sludge    Total
## 1       380686.2     0.00 71983.68 3771.40  68063.34 893360.9
## 2       401034.3  1516.12 71066.46 5066.35  65585.25 951141.6
## 3       362474.4 10891.73 78980.55 5323.75  79967.05 894435.6
## 4       347204.0  7964.83 56849.63 4575.75 141423.92 921344.5
## 5       334512.7 12756.90 82081.97 3402.03  83424.85 867126.5
## 6       310959.5 17946.71 75803.52 3616.26  72882.61 858042.2
```

```
tail(ADC_data)
```

```
##    Report.Year Report.Quarter     Ash Auto.Shredder.Waste
## 75        1999              3 1578.70            69300.25
## 76        1999              4 2718.22            63910.19
## 77        1998              1 2631.85            39181.17
## 78        1998              2  878.63            49391.25
```

```
## 79        1998              3 2457.00            35573.00
## 80        1998              4 2418.00            38495.89
##    Construction.and.Demolition.Waste Compost Contaminated.Sediment
## 75                          48321.13       0                  0.00
## 76                          62057.02     381                 16.50
## 77                           2693.48       0                  0.00
## 78                           6666.70       0                  2.74
## 79                          28278.30       0                 92.17
## 80                          29591.80       0                  0.00
##    Green.Material   Mixed   Other    Tires   Sludge    Total
## 75       349276.6    0.00 4695.69  1265.82 66864.38 541302.6
## 76       360153.2    0.00 6316.72  3307.48 69058.27 567918.6
## 77       191066.3 3907.20 1008.27 14802.71 43391.12 298682.1
## 78       279191.3 3602.22 3305.93 15394.54 92416.47 450849.8
## 79       299986.8    0.00 2706.53  2943.31 99312.34 471349.4
## 80       313452.3 4130.00 3767.93   733.71 57511.25 450100.9
```

```r
# tidy the data by gathering the type columns
ADC_gathered <- gather(ADC_data, "Type", "Quantity", Ash:Sludge) %>%
  select(-Total) # remove Total column

# save the tidy dataset
write.csv(ADC_data, row.names = FALSE,
          file = "../Processed_Data/CalRecycle_ADC_tidy_processed.csv")
```

3.2 Summary

```r
# generate summary data
ADC_summary_by_type <- ADC_gathered %>%
  group_by(Type) %>% # group the data by lakename
  filter(!is.na(Quantity)) %>% #remove the records when there are nas Quantity
  summarise(MeanQuarterlyQuantity = mean(Quantity),
            MinQuarterlyQuantity = min(Quantity),
            MaxQuarterlyQuantity = max(Quantity),
            sdQuarterlyQuantity = sd(Quantity),
            medianQuarterlyQuantity = median(Quantity))

ADC_summary_by_type_table <- kable(ADC_summary_by_type,
  col.names = c("Waste Type", "Mean Quarterly Quantity", "Min Quarterly Quantity",
  "Max Quarterly Quantity", "sd of Quarterly Quantity",
  "Median Quarterly Quantity")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",
                                      "full_width = F"), latex_options="scale_down") %>%
  row_spec(0, bold = T)
```

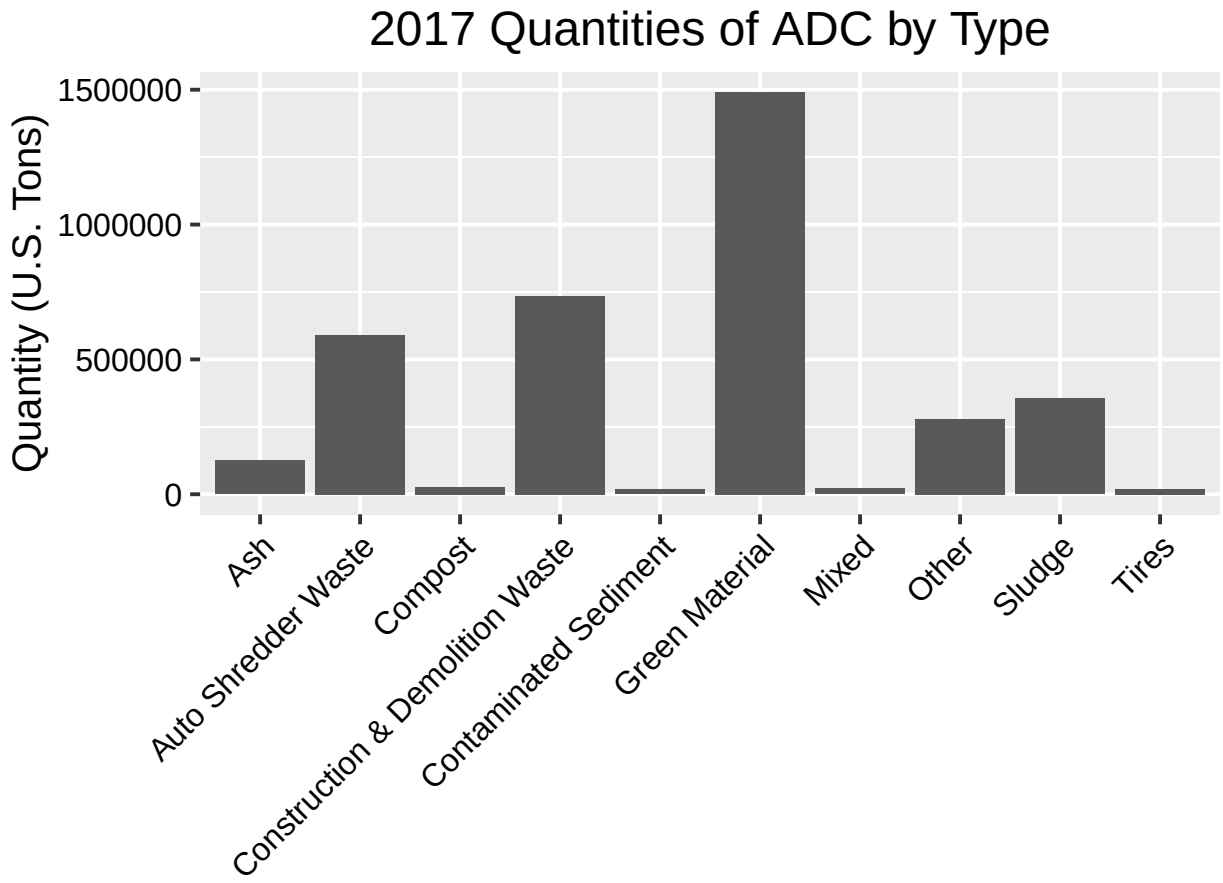
```
ADC_summary_by_type_table
```

Waste Type	Mean Quarterly Quantity	Min Quarterly Quantity	Max Quarterly Quantity	sd of Quarterly Quantity	Median Quarterly Quantity
Ash	11304.716	101.00	108208.23	17860.123	2675.035
Auto.Shredder.Waste	121864.560	35573.00	215857.61	39655.164	123356.730
Compost	3451.346	0.00	42215.62	6437.880	816.520
Construction.and.Demolition.Waste	122445.102	2693.48	281972.47	58940.701	125469.485
Contaminated.Sediment	11632.401	0.00	102850.25	20862.617	873.295
Green.Material	464016.026	191066.26	797463.76	137342.174	431369.340
Mixed	6929.398	0.00	99288.87	12855.562	3618.210
Other	45163.726	1008.27	149415.30	33034.798	47919.710
Sludge	73697.705	23812.68	147167.22	23765.135	72039.815
Tires	6544.532	733.71	29842.28	5735.144	4353.135

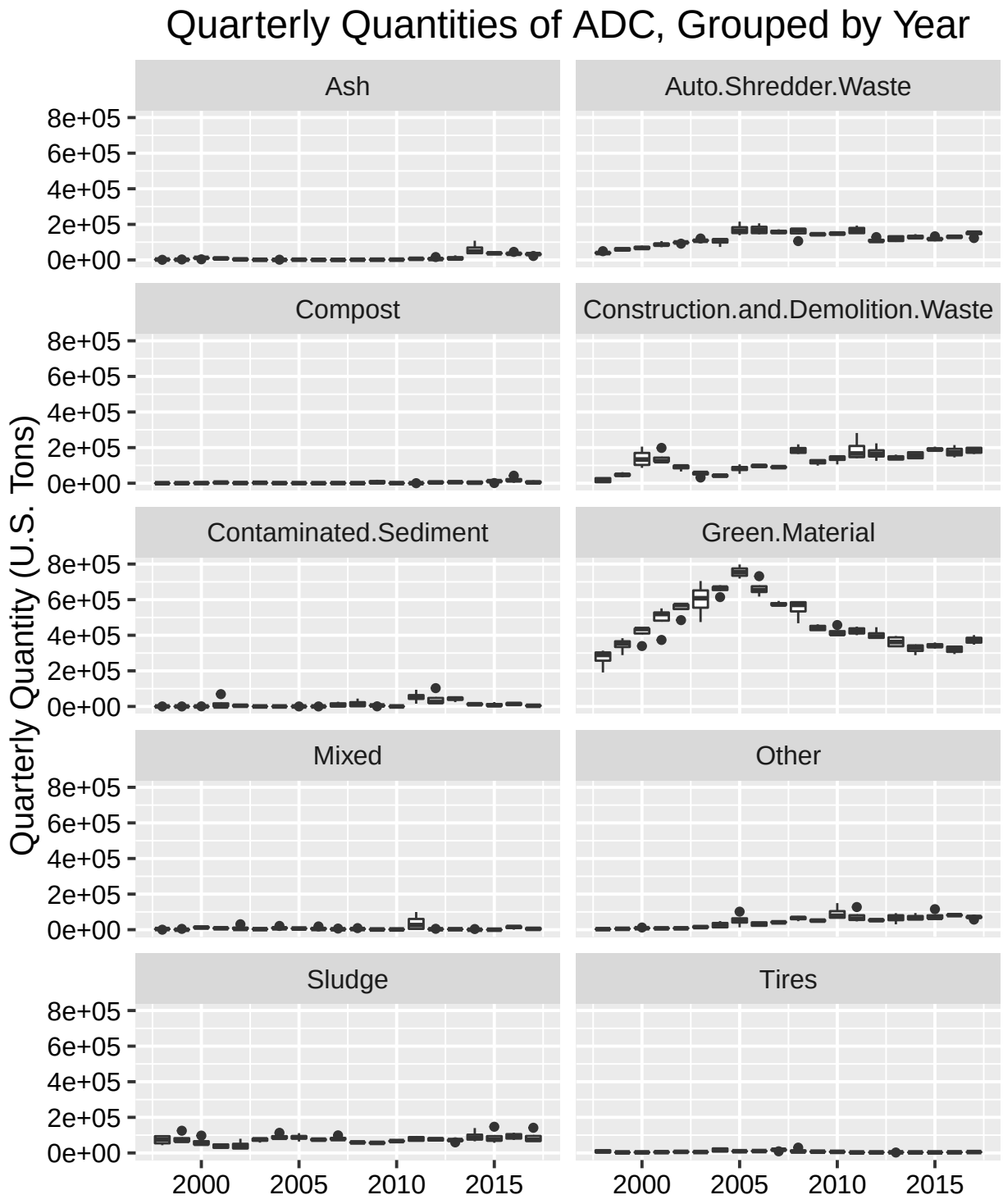
```
ADC_summary_by_year <- ADC_gathered %>%
  group_by(Report.Year) %>% # group the data by year
  filter(!is.na(Quantity)) %>% #remove the records when there are nas Quantity
  summarise(MeanQuarterlyQuantity = mean(Quantity),
            MinQuarterlyQuantity = min(Quantity),
            MaxQuarterlyQuantity = max(Quantity),
            sdQuarterlyQuantity = sd(Quantity),
            medianQuarterlyQuantity = median(Quantity))

ADC_summary_by_year_table <- kable(ADC_summary_by_year,
  col.names = c("Year", "Mean Quarterly Quantity", "Min Quarterly Quantity",
  "Max Quarterly Quantity", "sd of Quarterly Quantity",
  "Median Quarterly Quantity")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed",
                                      "full_width = F"), latex_options="scale_down") %>%
  row_spec(0, bold = T)
ADC_summary_by_year_table
```

Year	Mean Quarterly Quantity	Min Quarterly Quantity	Max Quarterly Quantity	sd of Quarterly Quantity	Median Quarterly Quantity
1998	41774.55	0.00	313452.3	82508.86	3685.075
1999	54320.11	0.00	383358.8	103282.44	4657.845
2000	72004.54	0.00	437691.8	124324.24	12219.725
2001	80768.84	0.00	550962.5	147577.35	10910.315
2002	81108.82	0.00	572016.4	162476.47	8921.990
2003	86182.39	0.00	704649.9	178905.18	8734.895
2004	94959.32	0.00	680872.4	193691.28	22845.585
2005	116741.86	0.00	797463.8	223003.44	11838.200
2006	105499.79	0.00	731872.6	196767.51	19307.055
2007	98051.49	0.00	592911.4	169159.05	33313.725
2008	104818.28	93.00	587091.4	164134.03	45992.905
2009	83490.23	264.00	461620.7	130916.23	31683.365
2010	87194.48	21.59	457554.9	125910.72	36318.670
2011	103442.44	90.18	448355.7	128119.69	60689.255
2012	87318.52	228.26	445240.2	120284.80	49674.030
2013	82700.26	31.08	395495.6	107155.27	52122.570
2014	85596.00	0.00	343124.0	97936.53	57852.960
2015	87924.03	0.00	359753.1	105114.41	49599.195
2016	88695.99	0.00	340147.2	96515.24	59193.530
2017	91507.06	0.00	401034.3	113972.21	47072.205

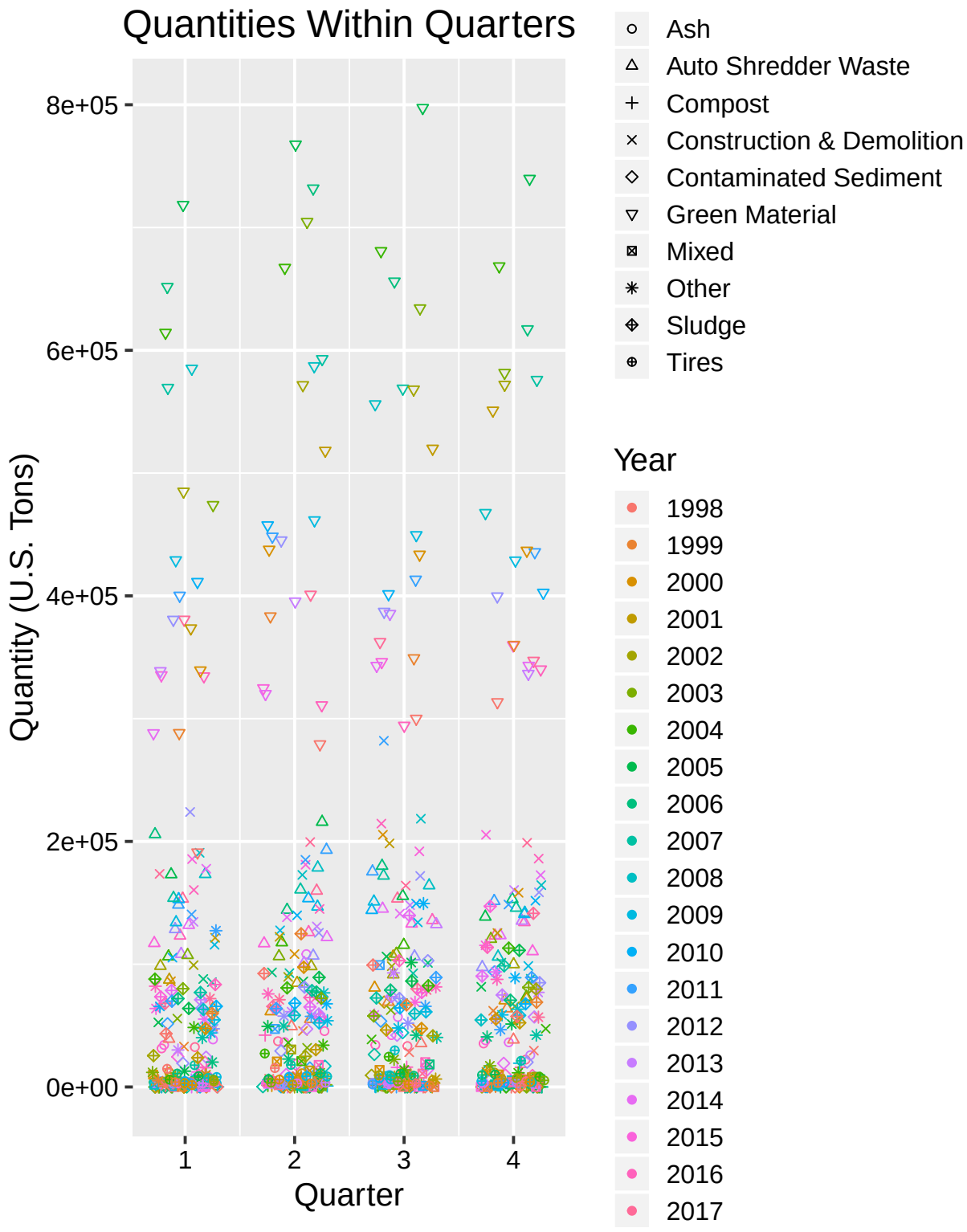
3.3 Exploratory Graphs



sko description graph1



sko description graph2



sko description graph 3

4 Analysis: Statistical Modeling & Data Visualization

4.1 Test 1: Difference Between Report Quarters - Analysis

Is there a significant difference in total ADC between report quarters? (i.e. 1, 2, 3, 4)

```
# create dataset with only total values, from 1995-2017
ADC_total_only <- ADC_raw %>%
  select(Report.Year, Report.Quarter, Total) # keep all columns except ADC Types

# convert column Report.Quarter into factor
class(ADC_total_only$Report.Quarter)
```

```
## [1] "integer"
```

```
ADC_total_only$Report.Quarter <- as.factor(ADC_total_only$Report.Quarter)

# save the dataset
write.csv(ADC_total_only, row.names = FALSE, file = "../Processed_Data/CalRecycle_ADC_to

# perform one-way ANOVA
# assumption #0: observations are independent (cannot be tested, but assumed to be ind

# test assumption #1: normality
# null hypothesis is that the dataset is normally distributed
shapiro.test(ADC_total_only$Total[ADC_total_only$Report.Quarter == 1]) # p-value = 0.03
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ADC_total_only$Total[ADC_total_only$Report.Quarter == 1]
## W = 0.90566, p-value = 0.03312
```

```
shapiro.test(ADC_total_only$Total[ADC_total_only$Report.Quarter == 2]) # p-value = 0.02
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ADC_total_only$Total[ADC_total_only$Report.Quarter == 2]
## W = 0.89774, p-value = 0.02271
```

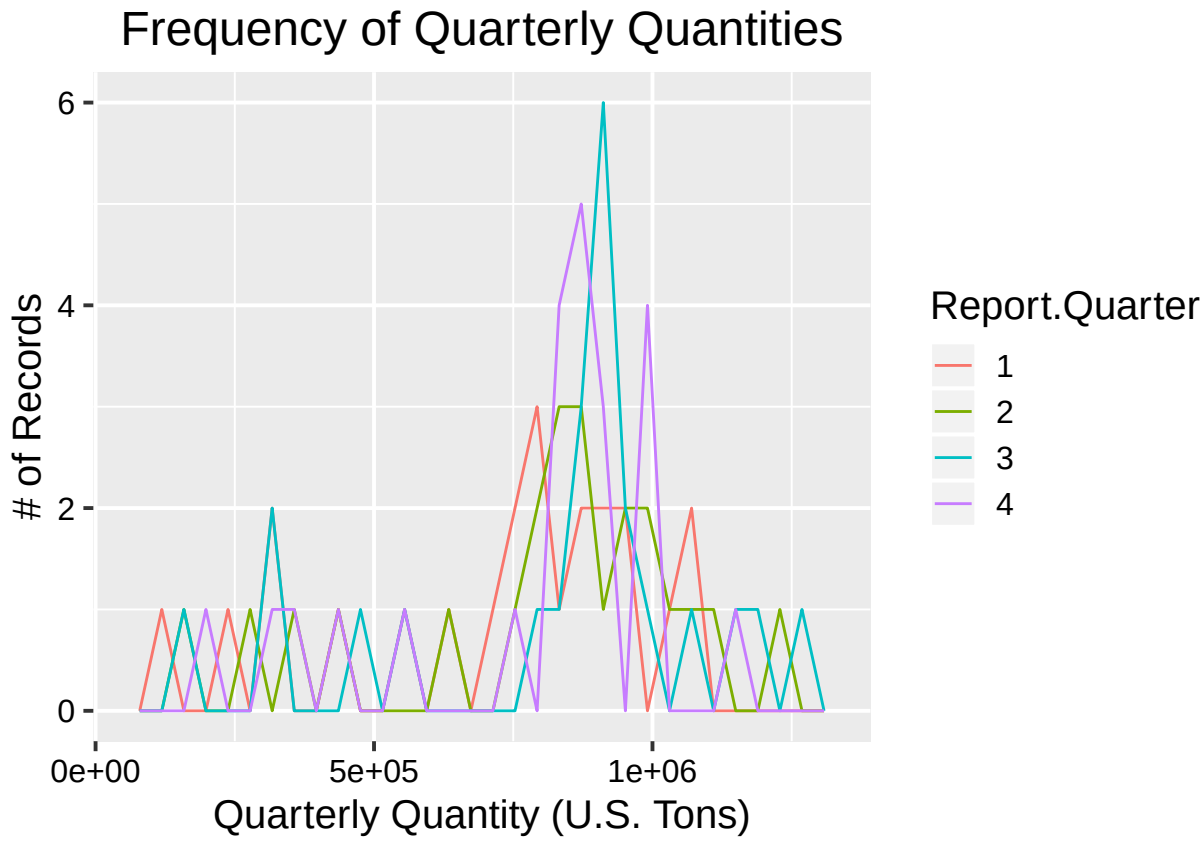
```
shapiro.test(ADC_total_only$Total[ADC_total_only$Report.Quarter == 3]) # p-value = 0.00
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ADC_total_only$Total[ADC_total_only$Report.Quarter == 3]
## W = 0.87982, p-value = 0.00993
```

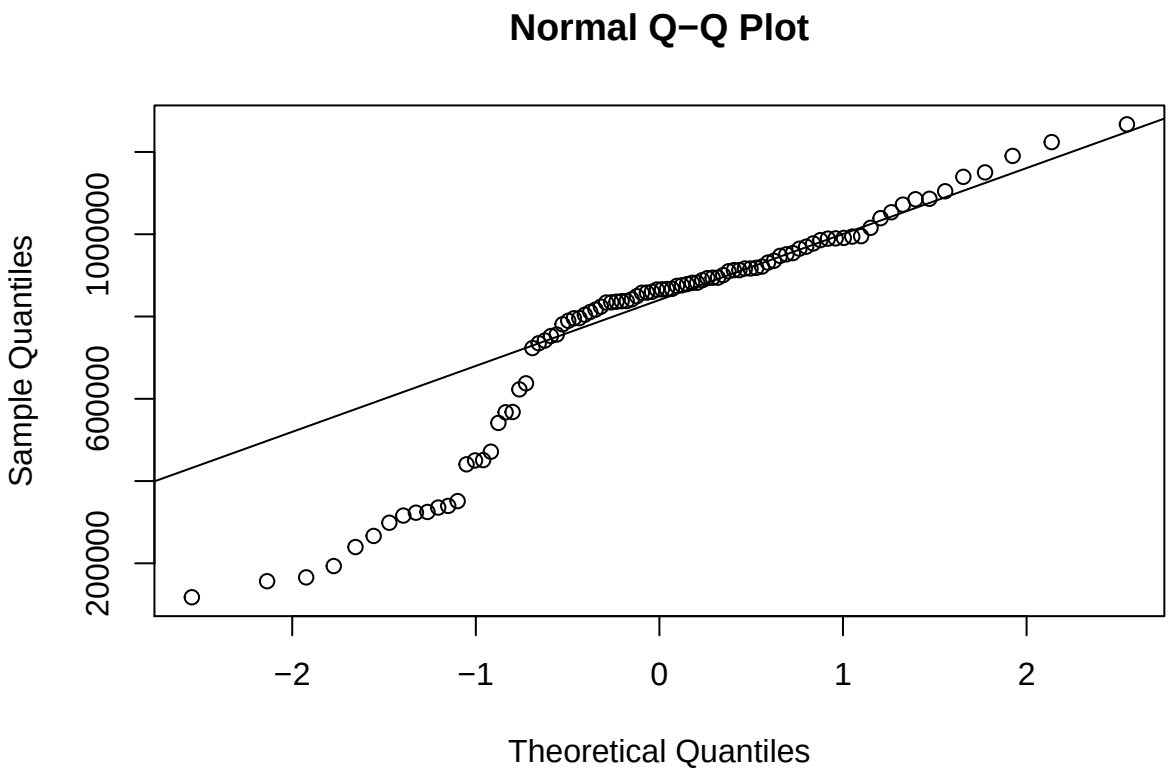
```r
shapiro.test(ADC_total_only$Total[ADC_total_only$Report.Quarter == 4]) # p-value = 0.00
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ADC_total_only$Total[ADC_total_only$Report.Quarter == 4]
## W = 0.83198, p-value = 0.001305
```

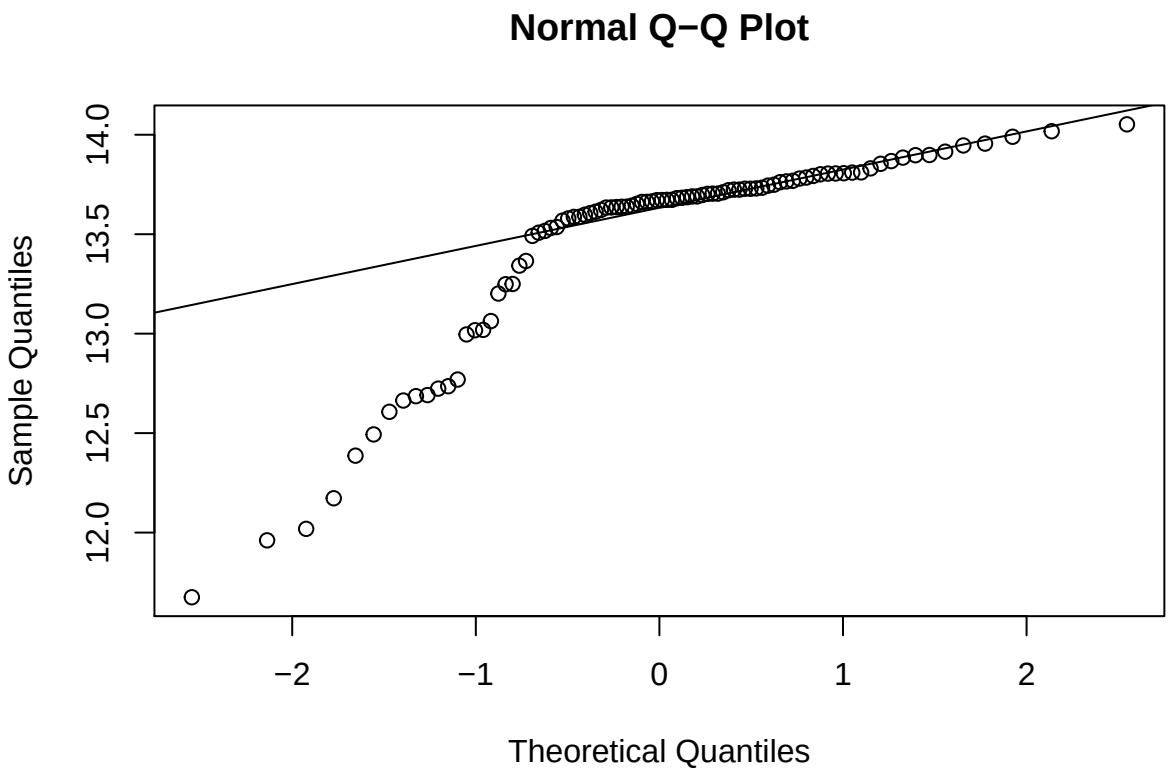
```r
ADC_freq_poly <- ggplot(ADC_total_only) +
  geom_freqpoly(aes(x = Total, color = Report.Quarter)) +
  xlab("Quarterly Quantity (U.S. Tons)") +
  ylab("# of Records") +
  ggtitle("Frequency of Quarterly Quantities")
print(ADC_freq_poly) # appears to be left skewed
```




```r
qqnorm(ADC_total_only$Total); qqline(ADC_total_only$Total) # does not match 1:1 ratio
```



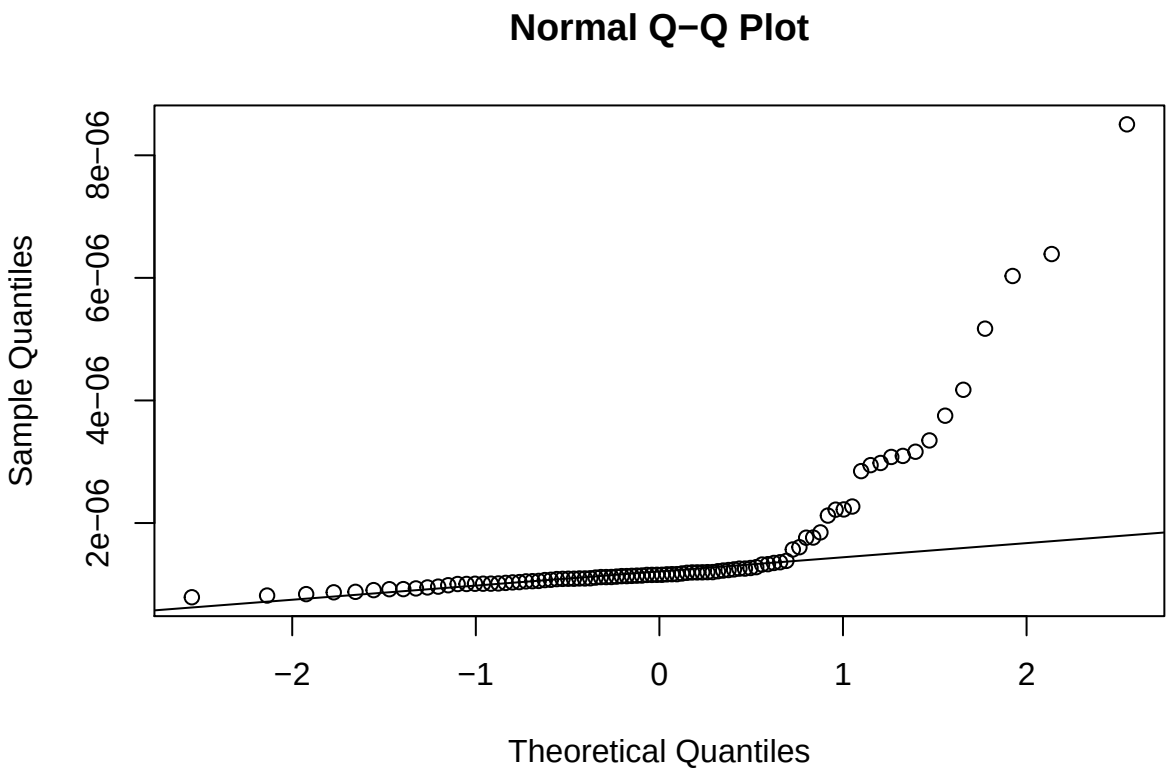

```
# Try to fix departure from normality with ln of Total. Result is not improved, so kee
ADC_LogTotal <- mutate(ADC_total_only, LogTotal = log(Total))
qqnorm(ADC_LogTotal$LogTotal); qqline(ADC_LogTotal$LogTotal)
```




```
bartlett.test(ADC_LogTotal$LogTotal ~ ADC_LogTotal$Report.Quarter)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  ADC_LogTotal$LogTotal by ADC_LogTotal$Report.Quarter
## Bartlett's K-squared = 1.1435, df = 3, p-value = 0.7666
```

```
# Try to fix departure from normality with 1/Total. Result is not improved, so keep no
ADC_InvTotal <- mutate(ADC_total_only, InvTotal = 1/Total)
qqnorm(ADC_InvTotal$InvTotal); qqline(ADC_InvTotal$InvTotal)
```




```
bartlett.test(ADC_InvTotal$InvTotal ~ ADC_InvTotal$Report.Quarter)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  ADC_InvTotal$InvTotal by ADC_InvTotal$Report.Quarter
## Bartlett's K-squared = 6.519, df = 3, p-value = 0.08892
```

```
# test assumption #2: equal variances among groups

# null hypothesis is that the variance is the same for the treatment groups
bartlett.test(ADC_total_only$Total ~ ADC_total_only$Report.Quarter) #p-value = 0.9308 #
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  ADC_total_only$Total by ADC_total_only$Report.Quarter
## Bartlett's K-squared = 0.44478, df = 3, p-value = 0.9308
```

```
# dataset is not normal, but does fulfill requirement for same variances. proceed with

# try non-parametric w/ post hoc, bc sample size is on the smaller end for parametric
ADC_quarter_kw <- kruskal.test(ADC_total_only$Total ~ ADC_total_only$Report.Quarter)
```

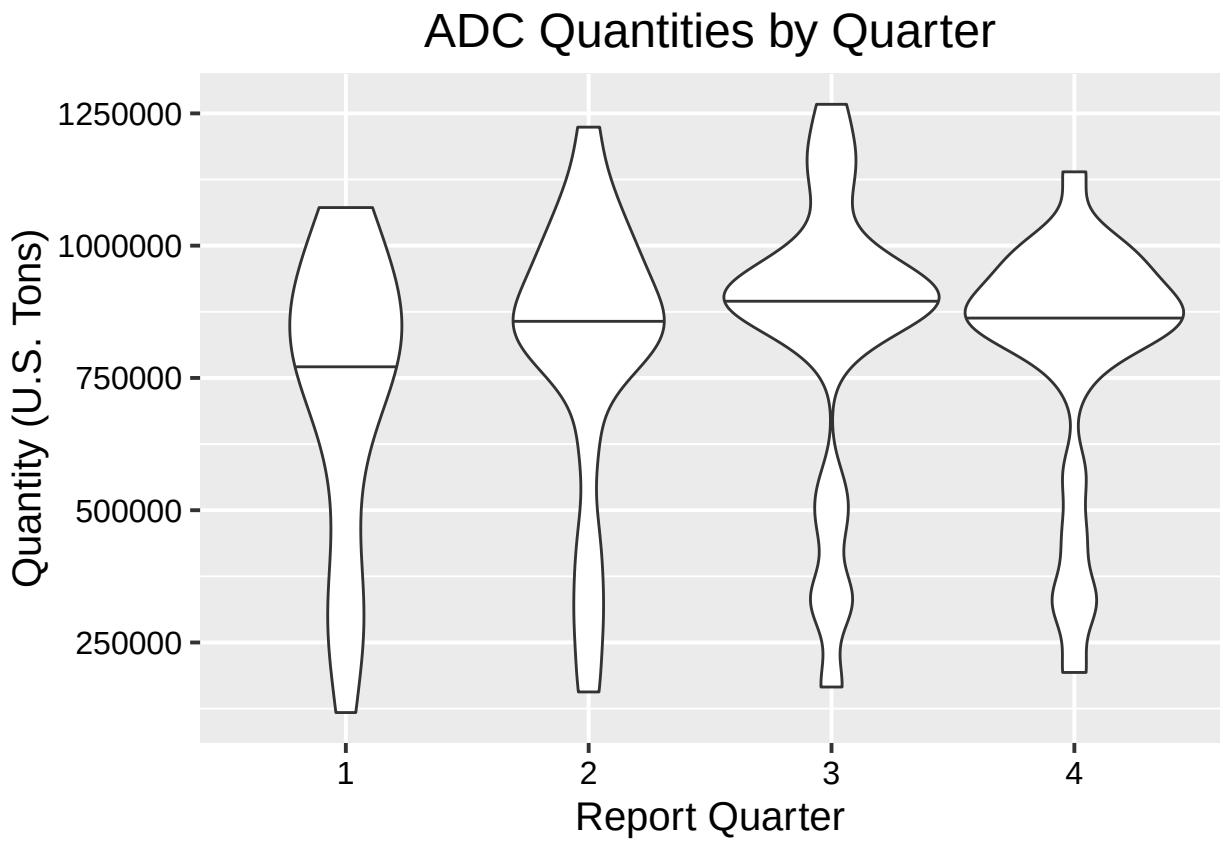
```
ADC_quarter_kw
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  ADC_total_only$Total by ADC_total_only$Report.Quarter
## Kruskal-Wallis chi-squared = 3.4581, df = 3, p-value = 0.3262
```

```
dunnTest(ADC_total_only$Total, ADC_total_only$Report.Quarter)
```

```
##   Comparison           Z    P.unadj     P.adj
## 1      1 - 2 -1.08778370 0.27669061 1.0000000
## 2      1 - 3 -1.84978446 0.06434462 0.3860677
## 3      2 - 3 -0.76200076 0.44605955 0.8921191
## 4      1 - 4 -1.00495753 0.31491730 1.0000000
## 5      2 - 4  0.08282617 0.93398976 0.9339898
## 6      3 - 4  0.84482693 0.39820748 1.0000000
```

4.1.1 Test 1: Difference Between Report Quarters - Result



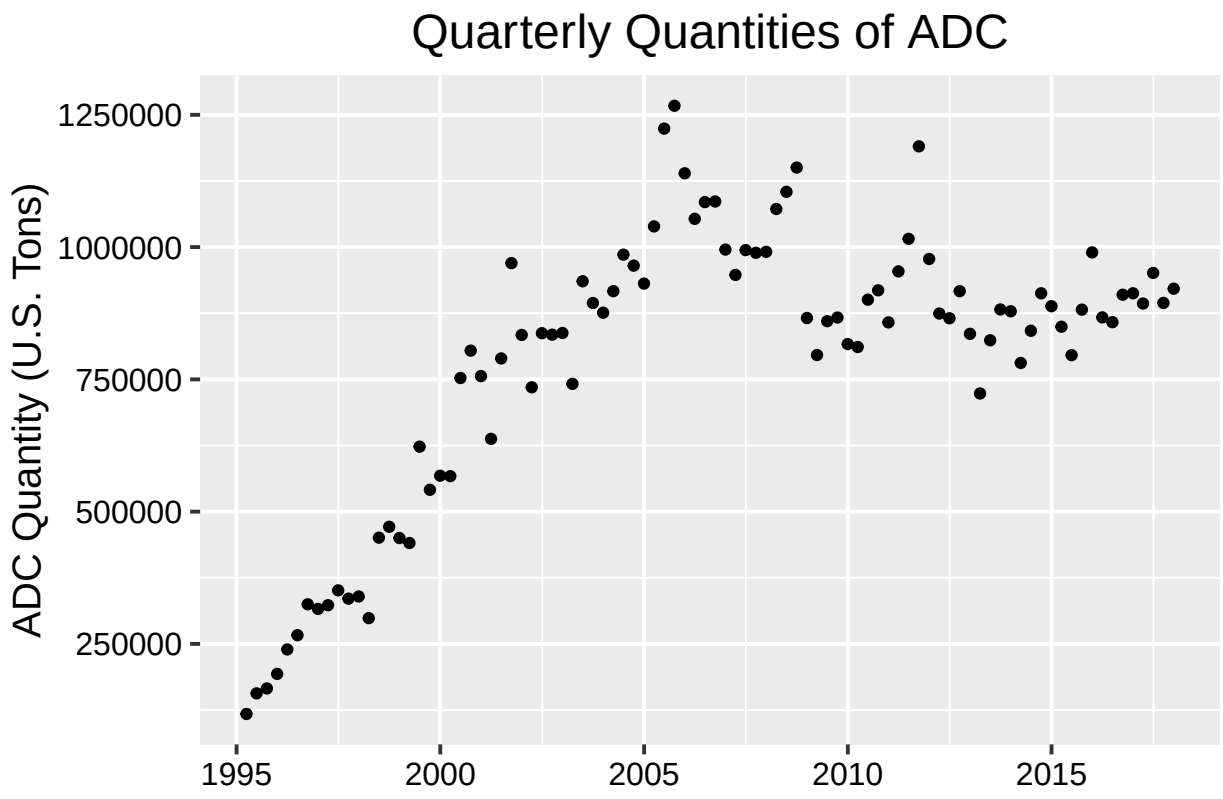
4.2 Test 2: Linear Model - Analysis

Can total annual ADC be represented with a linear model?

```
# assumptions for lm (independent observation, normal distribution, equal variances am

# create dates corresponding to year & quarter combination
# Q1: Mar 31
# Q2: Jun 30
# Q3: Sep 30
# Q4: Dec 31

# create dataframe of month-date
quarters_to_dates <- data.frame("Quarter" = as.factor(1:4), "Month.Date" = c('3-31', '6-

# create new dataframe with dates
ADC_fulldate <- ADC_total_only %>%
  inner_join(quarters_to_dates, by = c("Report.Quarter" =  "Quarter")) %>%
  unite('Quarter.End.Date', c(Report.Year, Month.Date), sep = "-", remove = FALSE)

ADC_fulldate$Quarter.End.Date <- as.Date(ADC_fulldate$Quarter.End.Date, "%Y-%m-%d")
class(ADC_fulldate$Quarter.End.Date)
```

```
## [1] "Date"
```

```
# create initial plot to visualize the data
ggplot(ADC_fulldate, aes(x = Quarter.End.Date, y = Total)) +
  geom_point() +
  xlab("") +
  ylab("ADC Quantity (U.S. Tons)") +
  ggtitle("Quarterly Quantities of ADC")
```




```
# create lm
ADC_date_lm <- lm(data = ADC_fulldate, Total ~ Quarter.End.Date)
ADC_date_lm # Total = 73.14*Quarter.End.Date - 190264.58
```

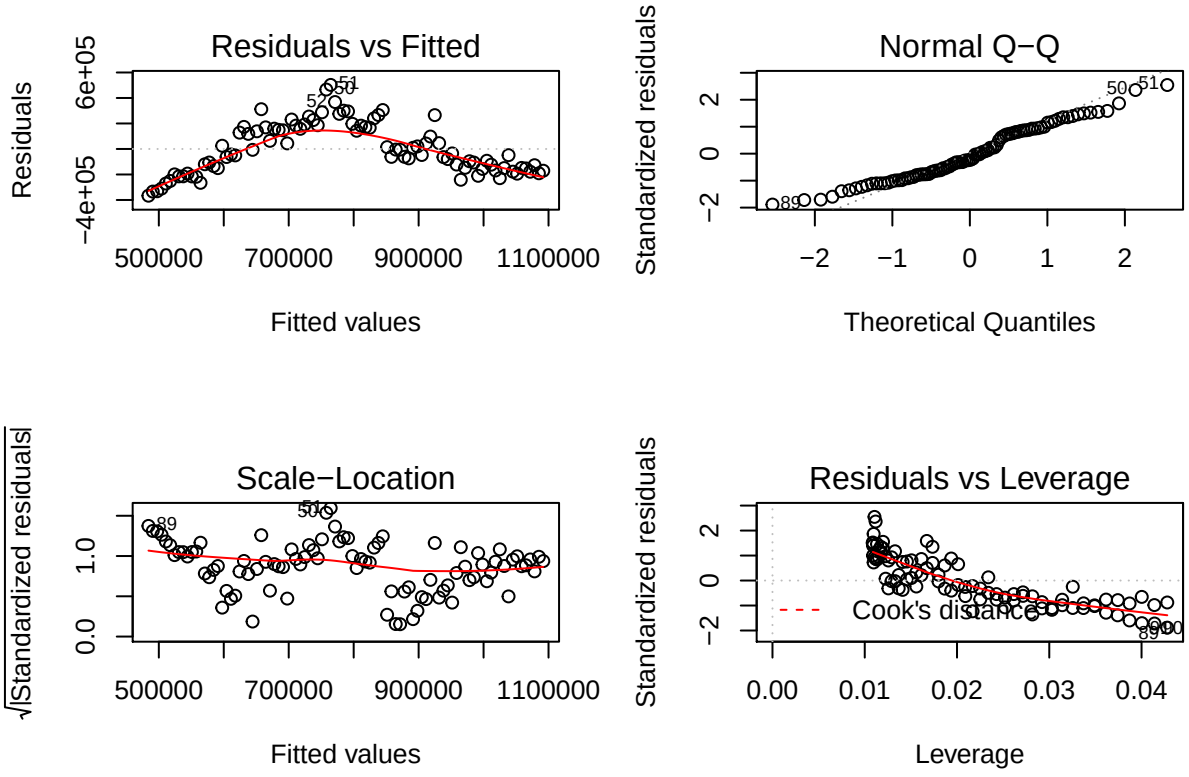
```
##
## Call:
## lm(formula = Total ~ Quarter.End.Date, data = ADC_fulldate)
##
## Coefficients:
##       (Intercept)  Quarter.End.Date
##        -190264.58             73.14
```

```
summary(ADC_date_lm) # Adjusted R-squared:  0.4433 (date explains 44.33% of variation
```

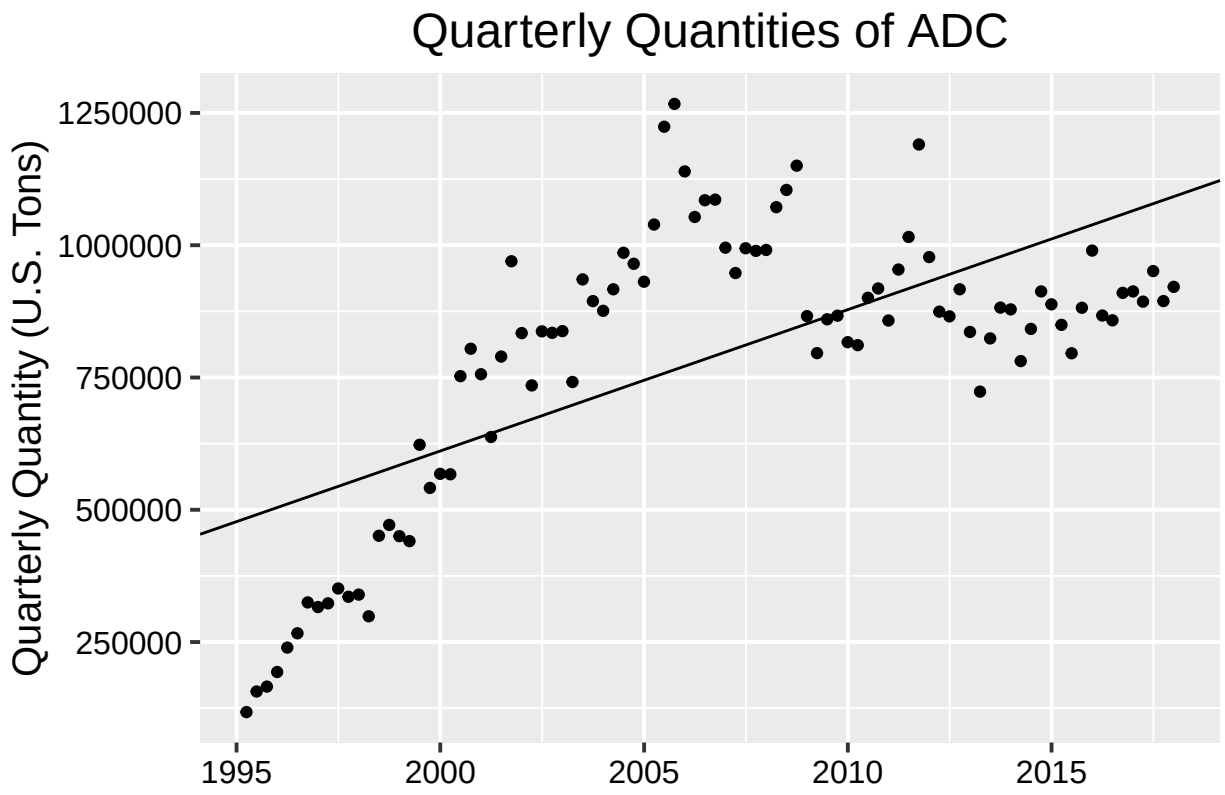
```
##
## Call:
## lm(formula = Total ~ Quarter.End.Date, data = ADC_fulldate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -366483 -153515  -45160  167108  502499
##
```

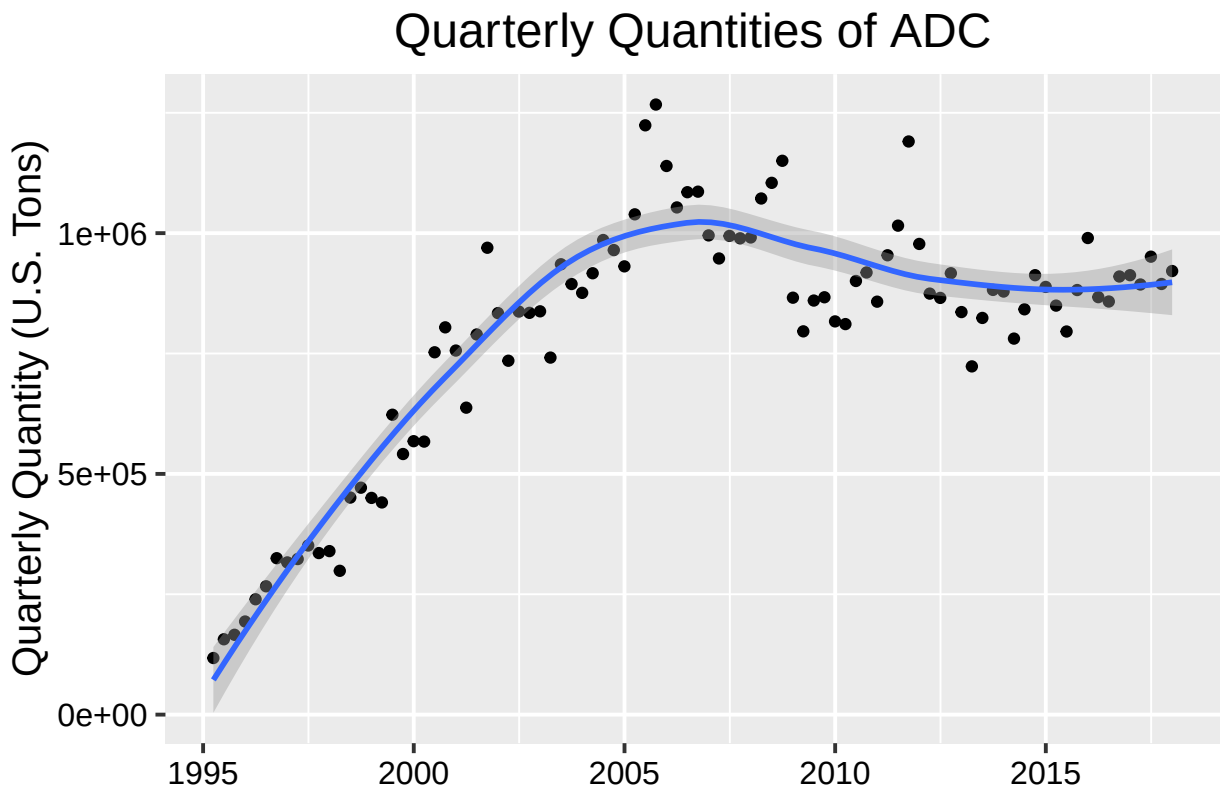
```
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.903e+05  1.160e+05   -1.64    0.104
## Quarter.End.Date  7.314e+01  8.534e+00    8.57 2.69e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 198500 on 90 degrees of freedom
## Multiple R-squared:  0.4494, Adjusted R-squared:  0.4433
## F-statistic: 73.45 on 1 and 90 DF,  p-value: 2.694e-13
```

```r
# check normality of residuals
par(mfrow=c(2,2))
plot(ADC_date_lm) # QQ of residuals looks relatively normal
```



4.2.1 Test 2: Linear Model - Result

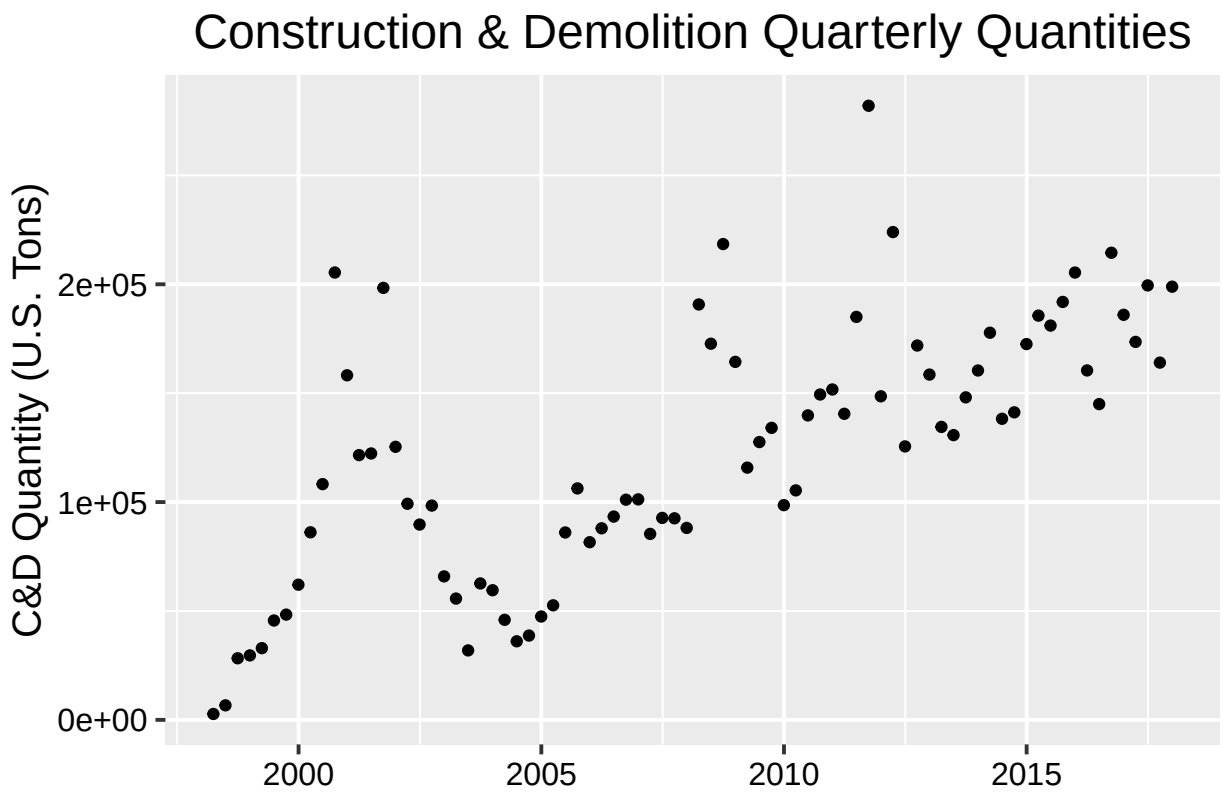




4.3 Test 3: Changepoint in Construction & Demolition - Analysis

Is there a changepoint in the Construction & Demolition quantities over time?

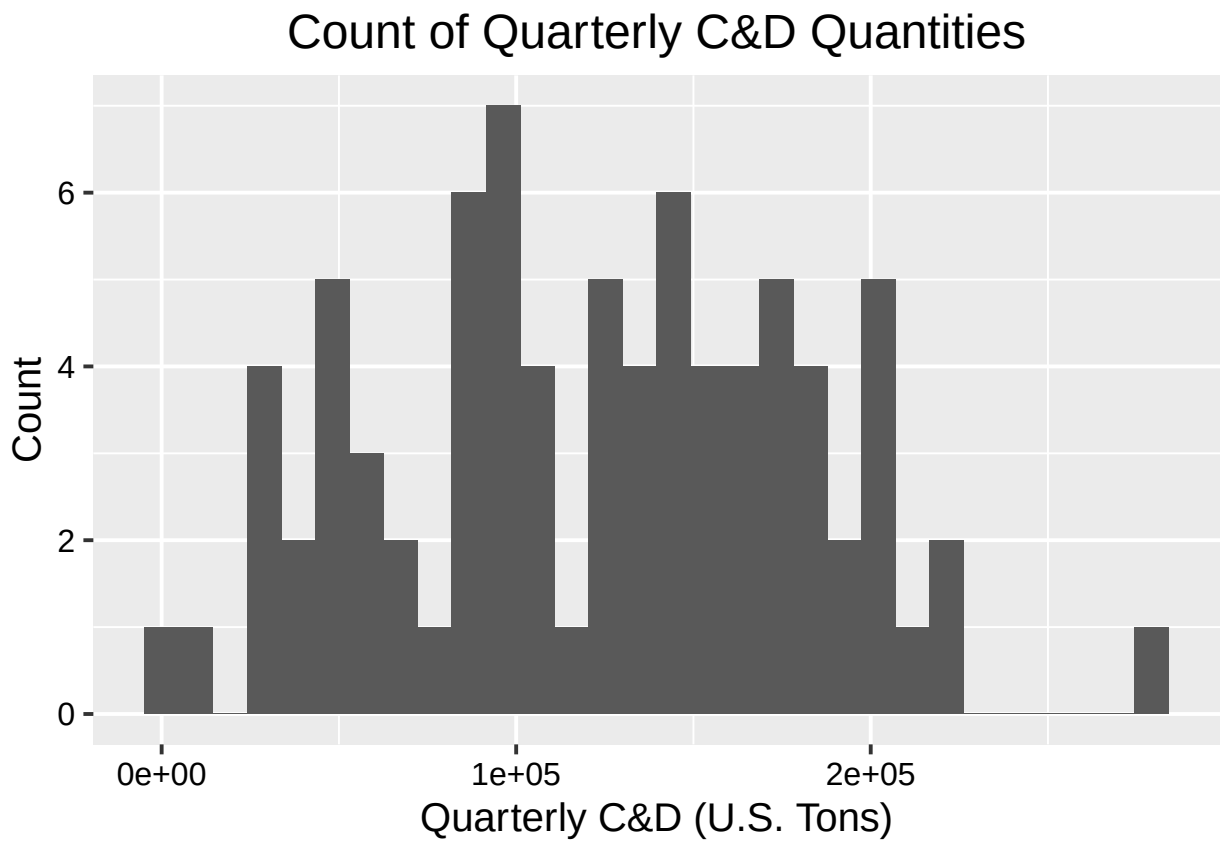
```
# create dataframe with dates
quarters_to_dates$Quarter <- as.integer(quarters_to_dates$Quarter)

CD_only <- ADC_data %>%
  select(Report.Year, Report.Quarter, Construction.and.Demolition.Waste) %>%
  inner_join(quarters_to_dates, by = c("Report.Quarter" =  "Quarter")) %>%
  unite('Quarter.End.Date', c(Report.Year, Month.Date), sep = "-") %>%
  select(-Report.Quarter)

CD_only$Quarter.End.Date <- as.Date(CD_only$Quarter.End.Date, '%Y-%m-%d') # format colu

# arrange data from oldest to newest
CD_only <- CD_only %>%
  arrange(Quarter.End.Date)

# create initial plot to visualize the data
ggplot(CD_only, aes(x = Quarter.End.Date, y = Construction.and.Demolition.Waste)) +
  geom_point() +
  xlab("") +
  ylab("C&D Quantity (U.S. Tons)") +
  ggtitle("Construction & Demolition Quarterly Quantities")
```



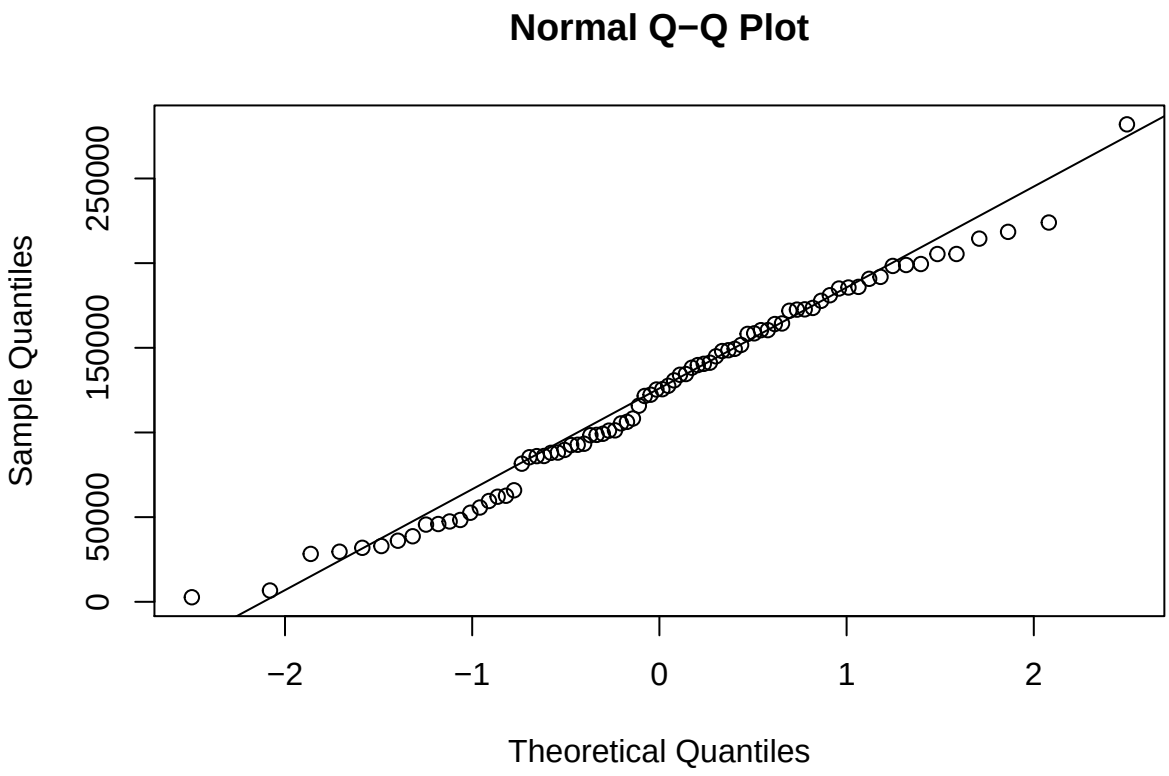

```
# check normality for C&D waste specifically
shapiro.test(CD_only$Construction.and.Demolition.Waste) # p-value = 0.4028, inferring t
```

```
##
##  Shapiro-Wilk normality test
##
## data:  CD_only$Construction.and.Demolition.Waste
## W = 0.9837, p-value = 0.4028
```

```
ggplot(CD_only) +
  geom_histogram(aes(x = Construction.and.Demolition.Waste)) +
  xlab("Quarterly C&D (U.S. Tons)") +
  ylab("Count") +
  ggtitle("Count of Quarterly C&D Quantities")
```




```
qqnorm(CD_only$Construction.and.Demolition.Waste); qqline(CD_only$Construction.and.Demol
```




```
# use Pettitt's test (nonparametric) to determine whether there is a shift in the cent
pettitt.test(CD_only$Construction.and.Demolition.Waste) # change point at time 40
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  CD_only$Construction.and.Demolition.Waste
## U* = 1396, p-value = 3.2e-10
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              40
```

```
# Run separate Mann-Kendall for each section
mk.test(CD_only$Construction.and.Demolition.Waste[1:40])
```

```
##
##  Mann-Kendall trend test
##
## data:  CD_only$Construction.and.Demolition.Waste[1:40]
## z = 1.736, n = 40, p-value = 0.08256
## alternative hypothesis: true S is not equal to 0
```

```
## sample estimates:
##            S         varS          tau
##  150.0000000 7366.6666667    0.1923077
```

```
mk.test(CD_only$Construction.and.Demolition.Waste[41:80])
```

```
##
##  Mann-Kendall trend test
##
## data:  CD_only$Construction.and.Demolition.Waste[41:80]
## z = 2.4817, n = 40, p-value = 0.01308
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##           S        varS         tau
##  214.000000 7366.666667    0.274359
```

```
# Is there a second change point?
pettitt.test(CD_only$Construction.and.Demolition.Waste[41:80])
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  CD_only$Construction.and.Demolition.Waste[41:80]
## U* = 203, p-value = 0.04614
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              27
```

```
# position 27, so 41+27 = change point at time 68

# Run separate Mann-Kendall for new section
mk.test(CD_only$Construction.and.Demolition.Waste[69:80]) # p-value = 0.9453, not likel
```

```
##
##  Mann-Kendall trend test
##
## data:  CD_only$Construction.and.Demolition.Waste[69:80]
## z = 0.068573, n = 12, p-value = 0.9453
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S         varS          tau
##   2.00000000 212.66666667   0.03030303
```

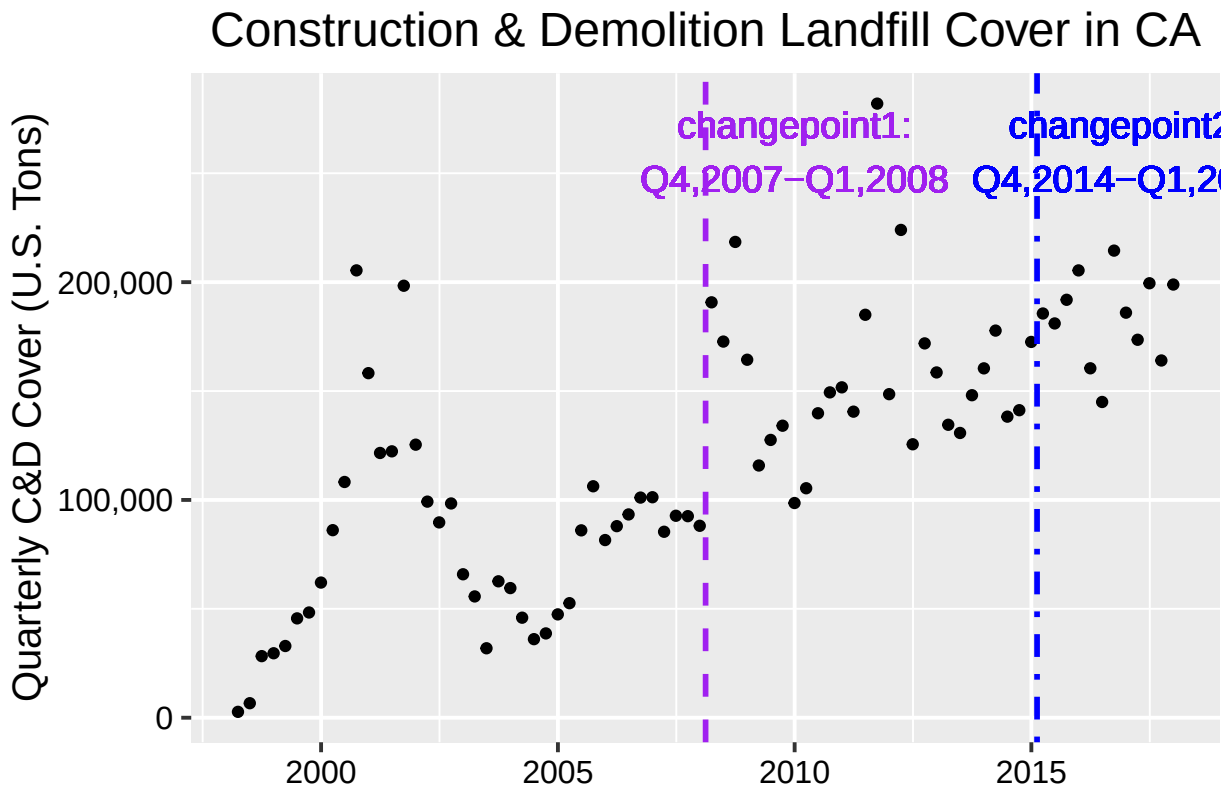
```
# Is there a third change point?
pettitt.test(CD_only$Construction.and.Demolition.Waste[69:80]) # p-value = p-value = 1.
```

```
##
```

```
##  Pettitt's test for single change-point detection
##
## data:  CD_only$Construction.and.Demolition.Waste[69:80]
## U* = 12, p-value = 1.261
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              6
```

```
# years corresponding to changepoints
changepoint1 <- CD_only$Quarter.End.Date[40] # between Q4 2007 & Q1 2008 = ~ 2008-02-14
changepoint2 <- CD_only$Quarter.End.Date[68] # between Q4 2014 & Q1 2015 = ~ 2015-02-14
```

4.3.1 Test 3: Changepoint in Construction & Demolition - Result



5 Summary and Conclusions