Trends in Alternative Daily Cover in California

https://github.com/sarahko7/ADC_Analysis

Sarah Ko

Abstract

Daily cover (DC) is the material, often soil, that is spread over garbage layers in a landfill. DC serves to reduce odor, slow infiltration by water, prevent garbage from being blown away, and deter pests like small mammals and birds. The state of California defines an alternative daily cover (ADC) as a material other than soil. This study uses the data from CalRecycle to investigate trends in state-wide use of alternative daily cover. When analyzing trends in reporting quarters, statistical tests found that quantities in each quarter are not significantly different from one another. In a second test, quarterly ADC quantities were modeled as a function of time - the linear model (Quarterly Quantity = 73.14*Date - 190264.58) was a moderate fit, explaining ~44% of the variation. However, the Loess regression appears to be a much better fit visually. For construction & demolition ADC specifically, tests detected 2 changepoints in the quantities over time - one around 2008-02-14, and another around 2015-02-14. It is notable though that these changepoints do not seem to be represented strongly in the visual depiction of the data. Findings from this analysis can be used to help rule makers and policymakers make more informed decisions regarding circular economy related standards and regulations.

Contents

List of Tables

List of Figures

1 Research Question and Rationale

Alternative daily cover (ADC) is a cover other than 'earthen' material that is spread over the active face of a municipal solid waste landfill at the end of each operating day. Certain regions in California can claim materials used for ADC as diverted from landfill[1], which is an important characteristic for certifications that are promoting circular economy concepts. As circular economy related rules and regulations change over time, the quantities of ADC may see changes as well. Construction & demolition (C&D) ADC is of particular interest because of a particular constituent - gypsum wallboard. Under the anerobic decomposition conditions of a landfill, this material forms hydrogen sulfide gas[2], which is not only highly toxic, but also corrosive to the methane collection equipment used in landfills.

This analysis has 3 main research questions:

1) For all ADC types combined, is there a significant difference between the quantities of ADC used in each reporting quarter?
2) For all ADC types combined, can the quarterly quantities be modeled as a function of time?
3) For construction & demolition ADC specifically, have there been changepoints in the quantities over time?

The first research question is an important component of understanding the trends of ADC over yearly seasons. This could signify an oversupply of ADC in a particular season (e.g. green material may be more plentiful in the summer), or an undersupply of earthen material. The second research question would be helpful in making predictions of future use of ADC. If the model predicts that the use of ADC is growing, this could be due to the influence of policies surrounding landfill diversion standards. The third research question looks at C&D ADC specifically - since this type of ADC is highly related to new and retrofit building construction, changepoints in this data could be related to trends in construction related sustainability standards. The dataset used to answer these questions is from the CalRecycle website - it includes quarterly information for ADC, split up by type. This allows for timeseries analyses, as well as analyses specific to certain waste types.

[1] https://www.calrecycle.ca.gov/lgcentral/basics/adcbasic
[2] http://www.newmoa.org/solidwaste/projects/gypsum.cfm

2 Dataset Information

The data used for this analysis was taken from CalRecycle - the database of the California Department of Resources Recycling and Recovery. This department is a part of the California Environmental Protection Agency. For more information about CalRecycle, see the 'About Us' section of the website[3].

The specific data used for this analysis is found on the page 'Statewide Alternative Daily Cover (ADC) by Material Type'[^4]. To extract the information, the data was exported to excel, then saved as a CSV file. The data for this analysis was extracted on 2009-04-06 (April 6, 2019).

ADC quantities are classified by 10 material types: Ash, Auto Shredder Waste, Compost, Construction & Demolition Waste, Contaminated Sediment, Green Material, Mixed, Other, Sludge, and Tires. A more detailed description of the data structure is found in Table 1. Classification of ADC into the full 10 categories was started in 1998. Prior to 1998, most of the ADC was categorized as 'Other'. The oldest datapoint in this record is Q1, 1995. The newest datapoint is Q4, 2017.

The quantities of ADC in this analysis are in units of U.S. tons, which is equivalent to 2000 lbs.

Table 1: Summary of Data Structure

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

[3] https://www.calrecycle.ca.gov/AboutUs/

[4]: https://www2.calrecycle.ca.gov/LGCentral/DisposalReporting/Statewide/ADCByMaterialType

3 Exploratory Data Analysis and Wrangling

3.1 Data Wrangling

The first part of wrangling the data was creating a dataset with only the years 1998-2017. The years before 1998 were removed because most of the ADC in that time was classified as 'Other'. If this data were included in analyses considering ADC types, the data from these years would make this category appear artificially large.

After reviewing the dataset, the second part of the wrangling process was to 'tidy' the data. The function 'gather' was used to rearrange the data under each type column into 2 new columns: ADC Type, and Quantity. This was done to facilitate visualization and statistical analyses later.

```r
# 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"
```

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

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

```r
# 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
```

```r
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

The following code is used to generate summary information about the data. This provides a general sense of the range and spread of the data. The 'group' and 'summarize' functions were used to generate this information.

Table 2 describes the summary statistics separated by type. Green material, construction & demolition, and auto shredder waste are seen to dominate the means. The sd column shows that green material has a much larger spread in its data as compared to the other types.

Table 3 describes the summary statistics separated by year. The mean quarterly quantities appear to peak between 2005-2008, and again in 2011.

```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)
```

Table 2: Summary statistics of ADC, separated by type

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

```r
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)
```

Table 3: Summary statistics of ADC, separated by year

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

Various graph types were used to illustrate the data exploration. These make it easy to get a general sense of the values in this dataset.

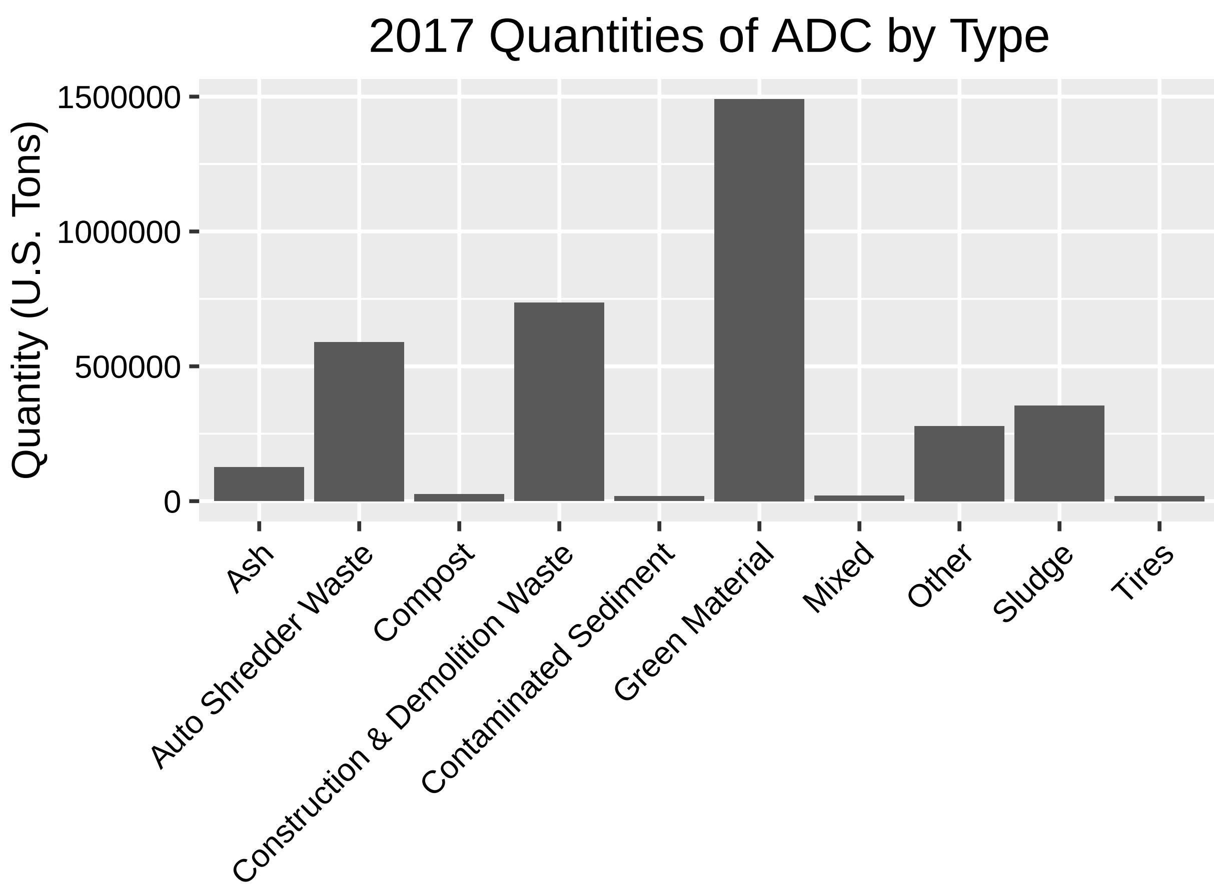


Figure 1: 2017 Quantities of ADC, separated by type

The geom_bar function was used to create Figure 1. This figure shows the 2017 ADC quantities per category. It is clear that waste categorized as compost, contaminated sediment, mixed, and tires makes up a very small part of the total ADC used.

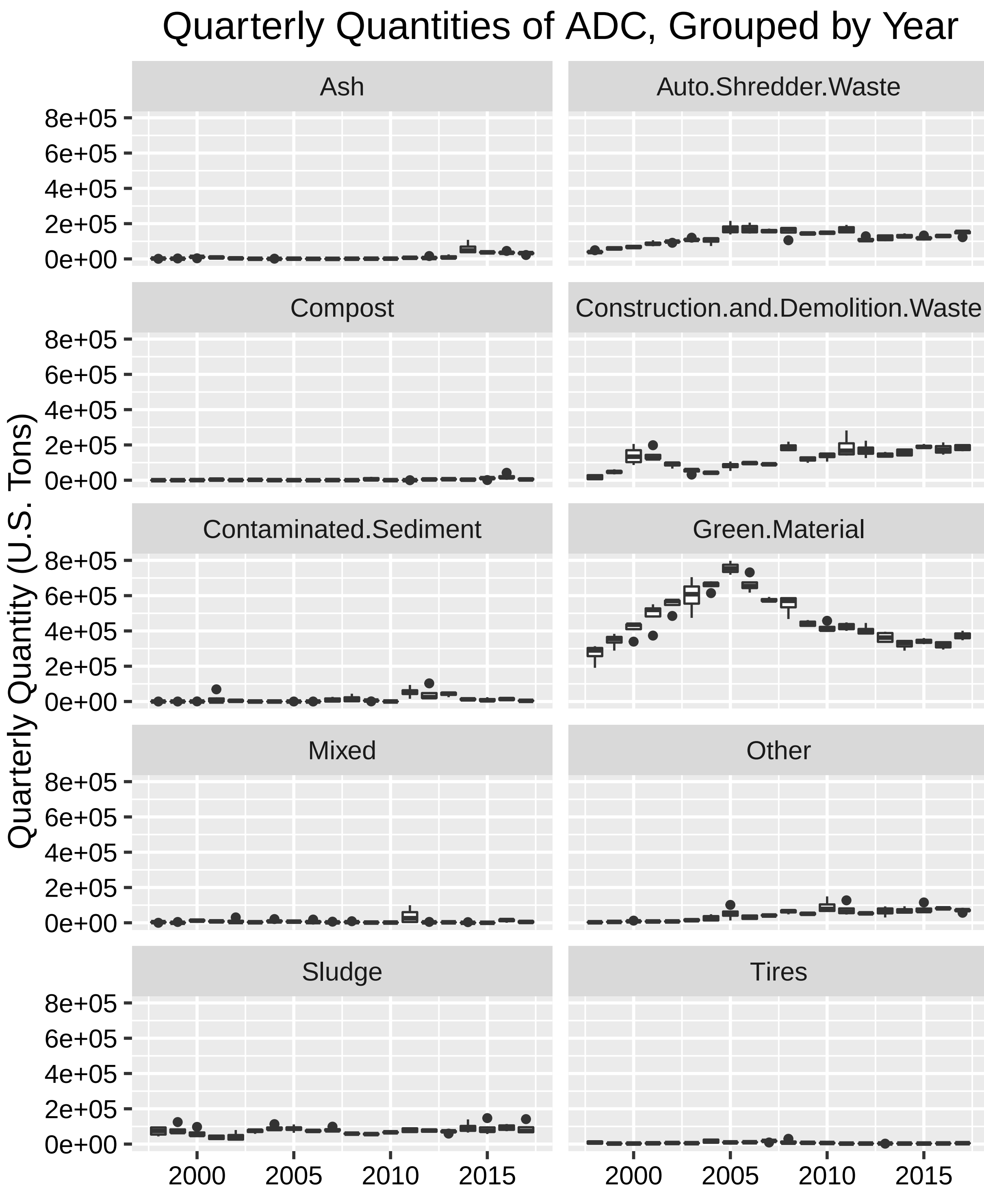


Figure 2: Quarterly quantities of ADC are grouped for each year. The data is displayed by

ADC type

The geom_boxplot function was used to create Figure 2, which the grouping function allowing the figure to display the data by year. The plot was then faceted by ADC type to view the fluctuations of the different ADC types over time. Auto shredder waste, and construction & demolition waste are seen to have experienced a gradual increase. The box plots illustrate the spread of the quarterly data within each year. Construction & demolition waste, and green material are notable in that they have a few years with very large spreads - C&D had large spreads in 2000 and 2010, and green material had large spreads from 1998-2005.

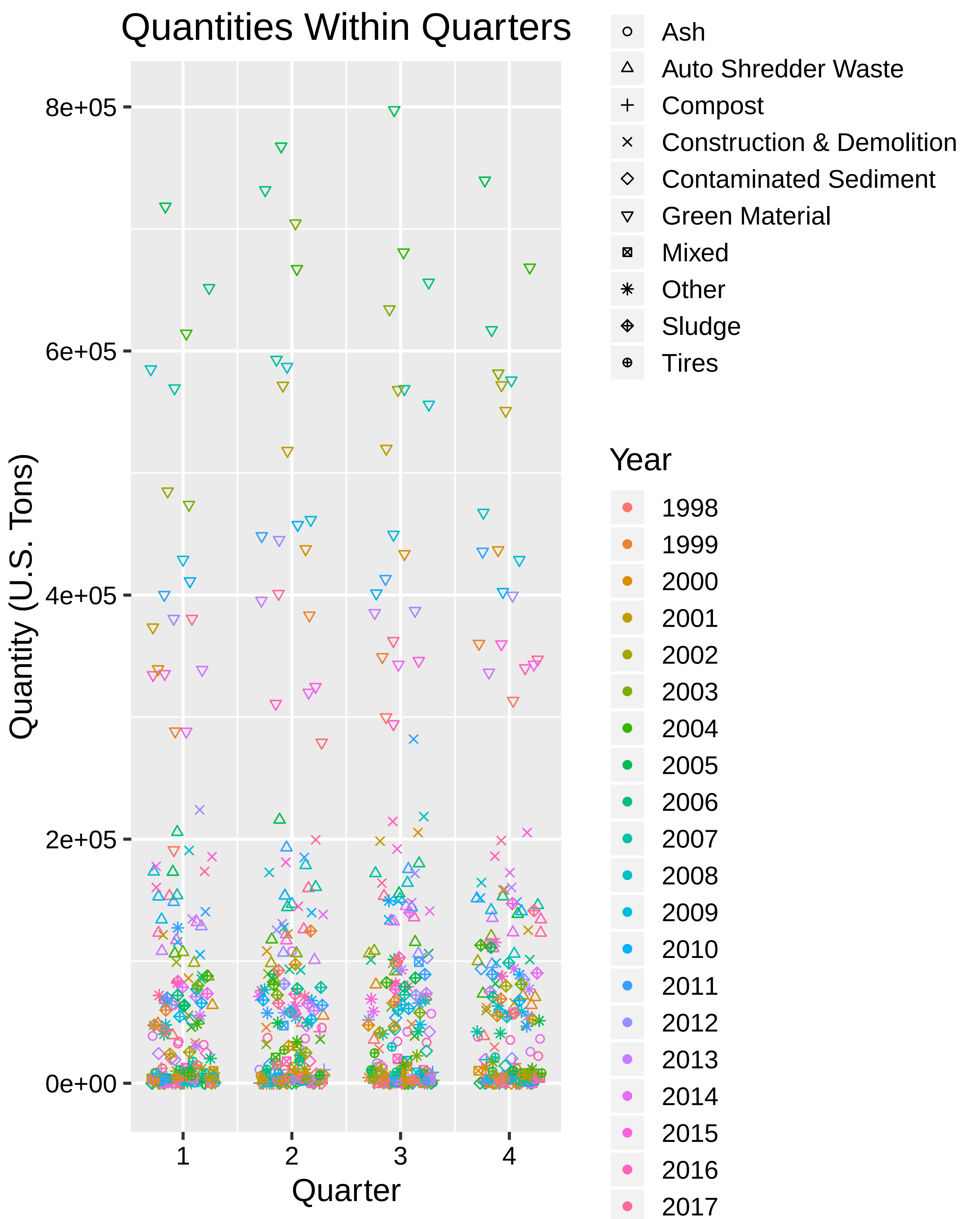


Figure 3: Quantities of ADC are grouped by the reporting quarter. The report year is classified by color, and the ADC type is classified by icon shape

gg_plot was used to create Figure 3, which shows the quarterly quantities of ADC split up by quarter. The jitter function was used to prevent some overlap, allowing more of the points to be seen. The characteristics of ADC type and year were illustrated the shape and color of the point, respectively.

4 Analysis: Statistical Modeling & Data Visualization

4.1 Test 1: Difference Between Report Quarters - Analysis

To assess trends in ADC corresponding to the yearly seasons, statistical analysis was used to check for significant differences in total ADC between report quarters 1, 2, 3, and 4. Since this analysis is on total ADC, a new dataset was created that included data from years 1995-2017. The quantity analyzed was the total (i.e. the sum of all the types, per year)

The test performed was a one-way ANOVA. The assumptions of a one-way ANOVA are: 1) the observations are independent 2) the groups are normally distributed 3) the variances among the groups are equal

Assumption 1 for independent observations cannot be checked from the dataset, and was assumed to be true.

Assumption 2 for normal distribution was checked for each quarter using the Shapiro Wilks test. The H0 for this test is that of normality, and the p values for each of the groups was $<$ 0.05 therefore these groups were deemed not normally distributed.

A frequency polygon was used to view the distribution of each quarter, which showed the data to be left skewed. A Q-Q plot was also used to check normality, and found that the data does not match the 1:1 ratio well. In an attempt to fix the departure from normality, a ln transformation and an inverse transformation were used, but neither was successful in making the data match a normal distribution.

Assumption 3 for homogeneity of variances was checked using the bartlett test. The H0 for this test is that the variance is equal among the groups. The p values were all $>$ 0.05 therefore the variances were deemed equal.

Since the data was deemed not normal, a Kruskal-Wallis, a non-parametric test, was used instead of the one-way ANOVA. The Dunn test, a post-hoc test, was also performed. The p values seen in the post-hoc test were all much greater than 0.05, therefore the means of these groups were found to not be significantly different from each other.

```r
# 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"
```

```r
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_totalsonly_processed.csv")
```

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

# 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])
```

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

```r
# p-value = 0.03312
shapiro.test(ADC_total_only$Total[ADC_total_only$Report.Quarter == 2])
```

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

```r
# p-value = 0.02271
shapiro.test(ADC_total_only$Total[ADC_total_only$Report.Quarter == 3])
```

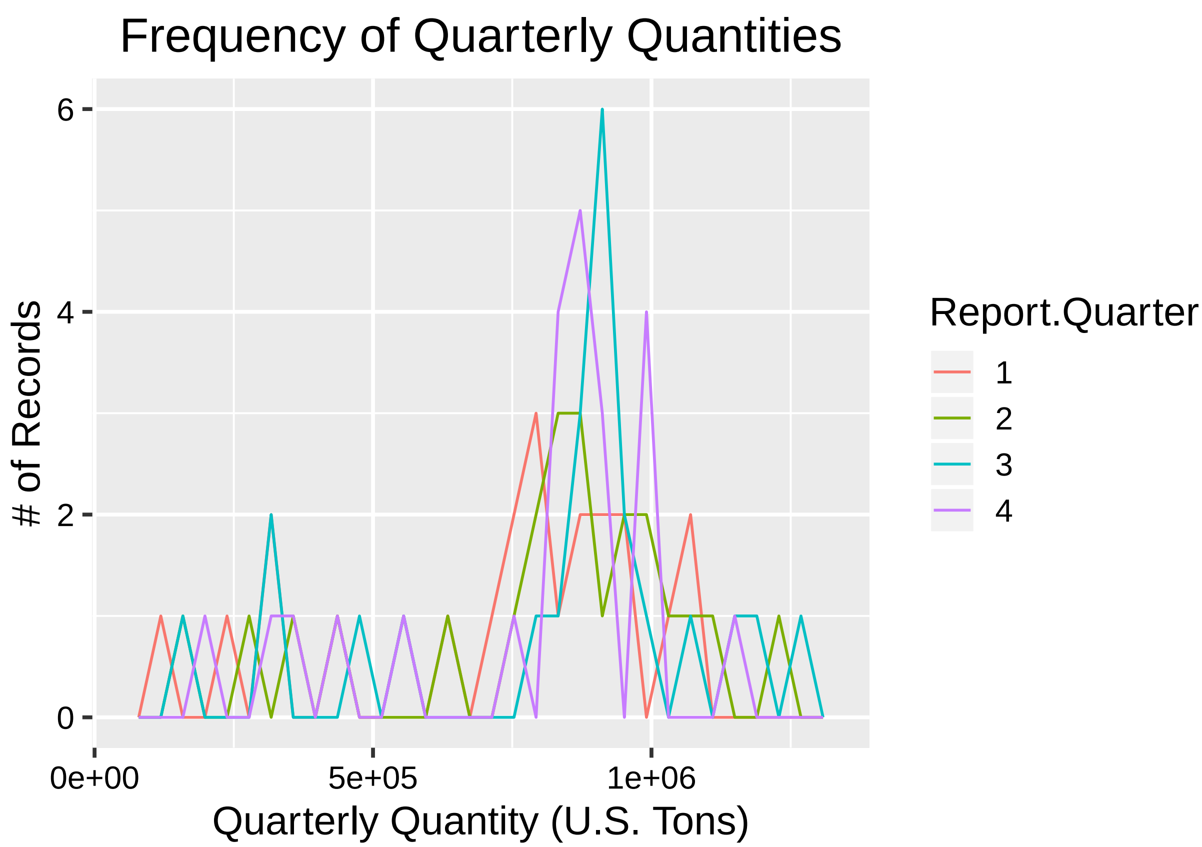
```
##
##  Shapiro-Wilk normality test
##
## data:  ADC_total_only$Total[ADC_total_only$Report.Quarter == 3]
## W = 0.87982, p-value = 0.00993
```

```r
# p-value = 0.00993
shapiro.test(ADC_total_only$Total[ADC_total_only$Report.Quarter == 4])
```

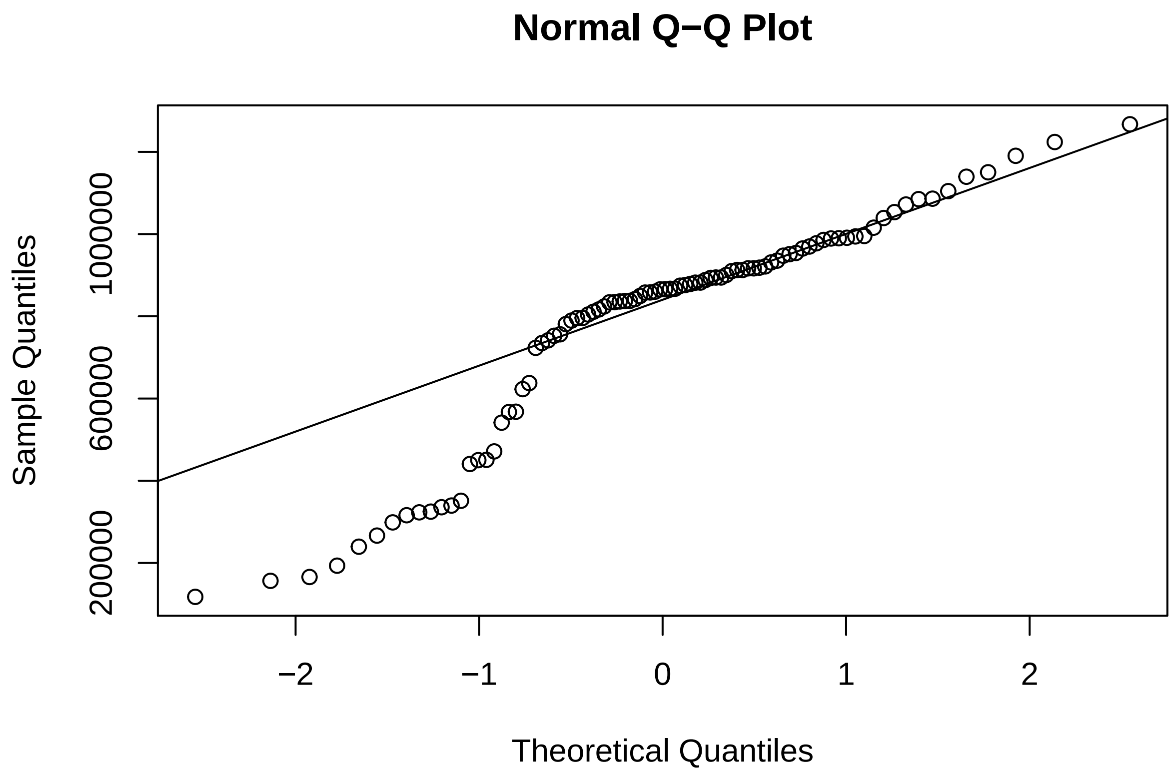
```
##
##  Shapiro-Wilk normality test
##
## data:  ADC_total_only$Total[ADC_total_only$Report.Quarter == 4]
## W = 0.83198, p-value = 0.001305
```

```r
# p-value = 0.001305

ADC_freq_poly <- ggplot(ADC_total_only) +
  geom_freqpoly(aes(x = Total, color = Report.Quarter)) +
  xlab("Quarterly Quantity (U.S. Tons)") +
  ylab("# of Records") +
```

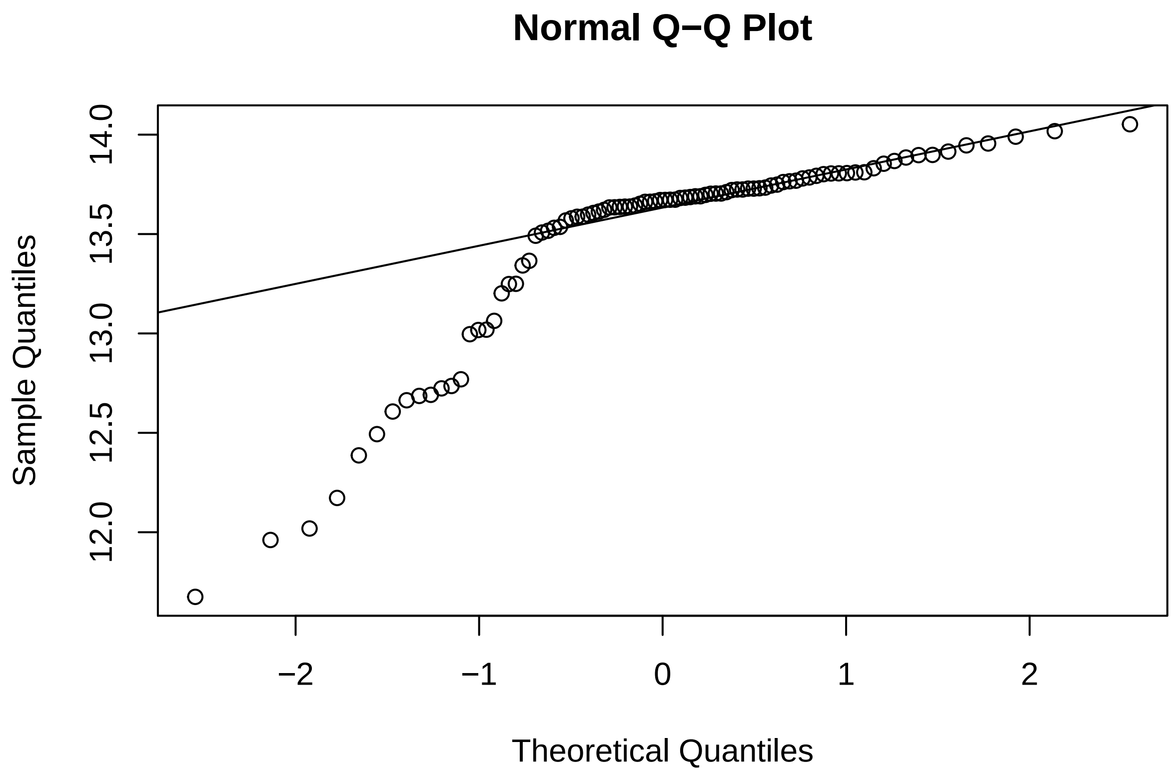
```
  ggtitle("Frequency of Quarterly Quantities")
print(ADC_freq_poly) # appears to be left skewed
```




```
qqnorm(ADC_total_only$Total); qqline(ADC_total_only$Total)
```



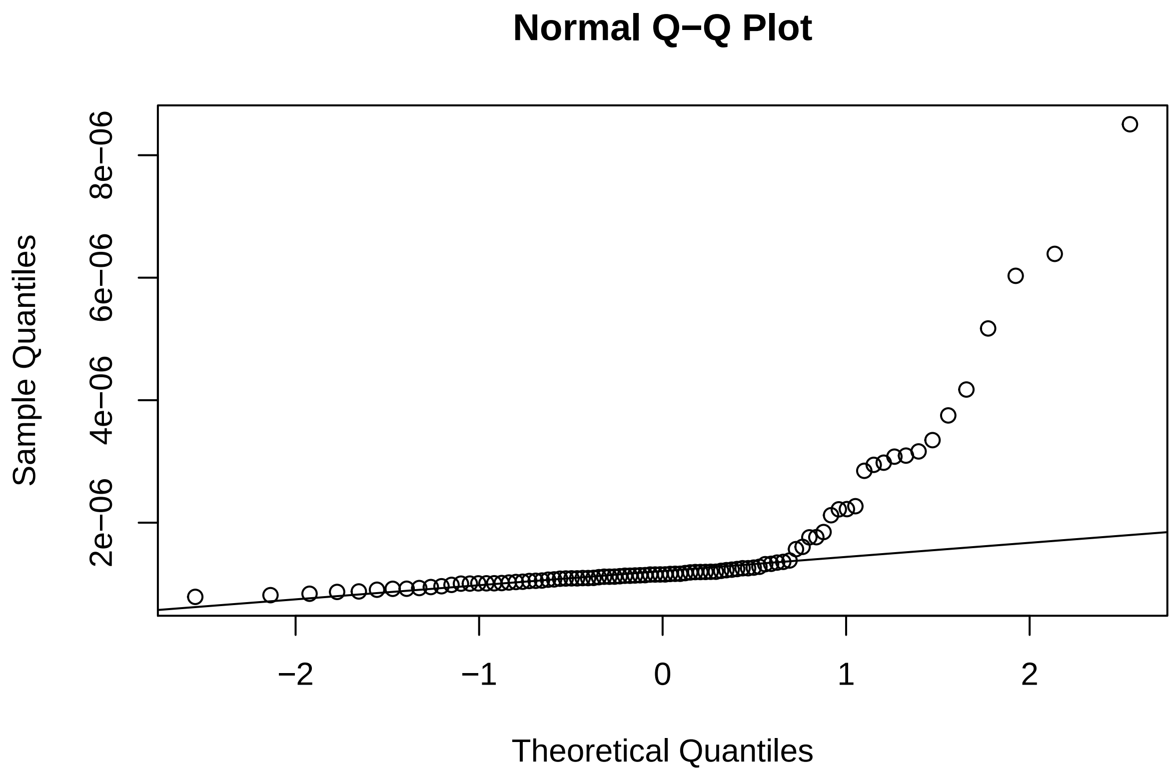

```
# does not match 1:1 ratio

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




```r
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
```

```r
# Try to fix departure from normality with 1/Total.
#Result is not improved, so keep non-transformed data
ADC_InvTotal <- mutate(ADC_total_only, InvTotal = 1/Total)
qqnorm(ADC_InvTotal$InvTotal); qqline(ADC_InvTotal$InvTotal)
```




```r
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
```

```r
# 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)
```

```
##
##  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
```

```r
#p-value = 0.9308 # df = 3 (statistical power is very low)

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

```r
# 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
```

```r
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: Result

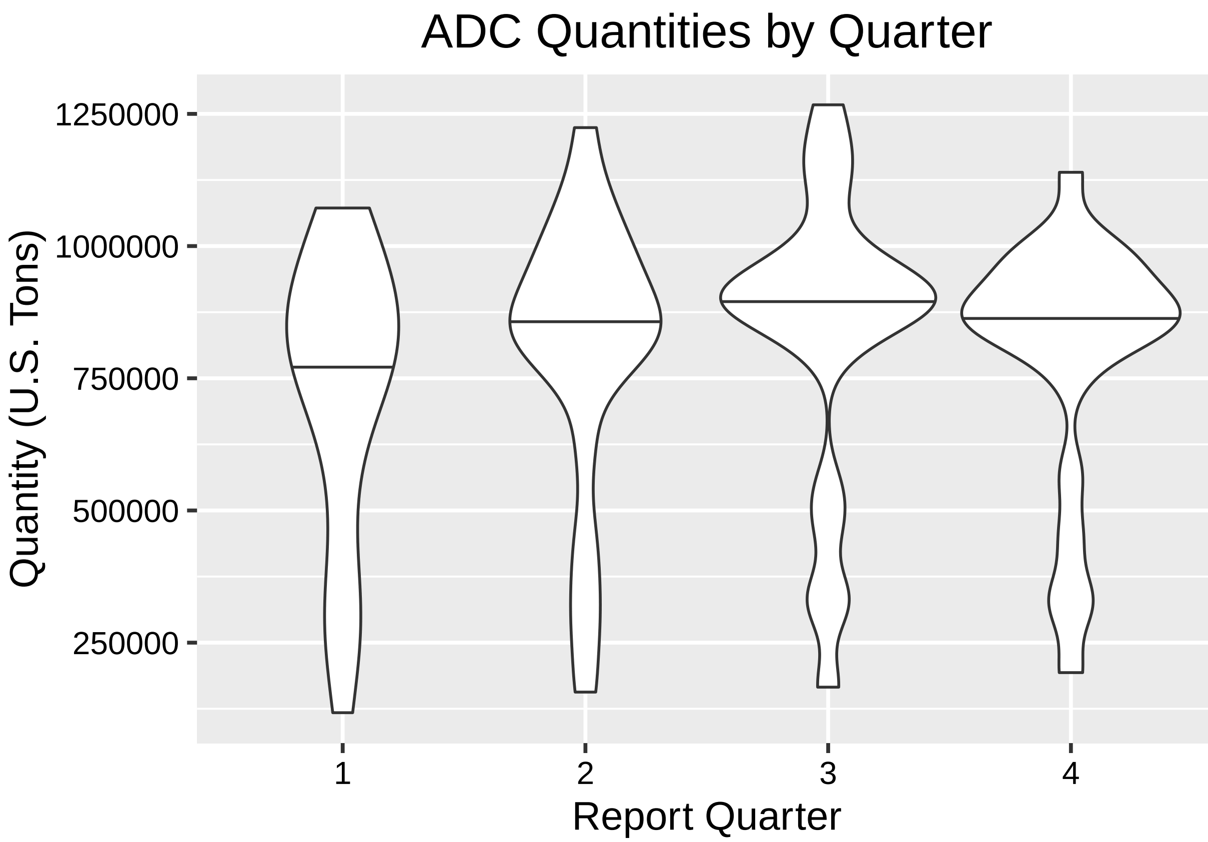


Figure 4: Quantities of ADC are grouped by quarter. The height of the bar described the spread of the data, a the width of the bar describes the number of records at each quantity

Figure 4 illustrates the numerical findings of the dunn test - that the means of the quarters are not significantly different from one another. The means of these groups is illustrated by the black horizonal lines, which visually appear to be at similar quantities.

4.2 Test 2: Linear Model - Analysis

The next test is to create a linear model of the quarterly values over time, and plot the model over the data points to visualize the fit. If the model is a good fit, the equation could be used to predict future quantities of ADC. Since this model does not evaluate the ADC quantities split up by type, the analysis is on the data from years 1995-2017. The quantity analyzed was the total (i.e. the sum of all the types, per quarter). To prepare this data for analysis, the quarter+year combinations were transformed into dates. The following dates were used for each quarter: Q1: Mar 31 Q2: Jun 30 Q3: Sep 30 Q4: Dec 31 These dates represent the end of each report quarter.

The assumptions for using a linear model are the same as that of the one-way ANOVA: 1) the observations are independent 2) the groups are normally distributed 3) the variances among the groups are equal

These were checked in Test 1. The data was not seen to be normal, but the group variances were found homogenous, so a lm is attempted.

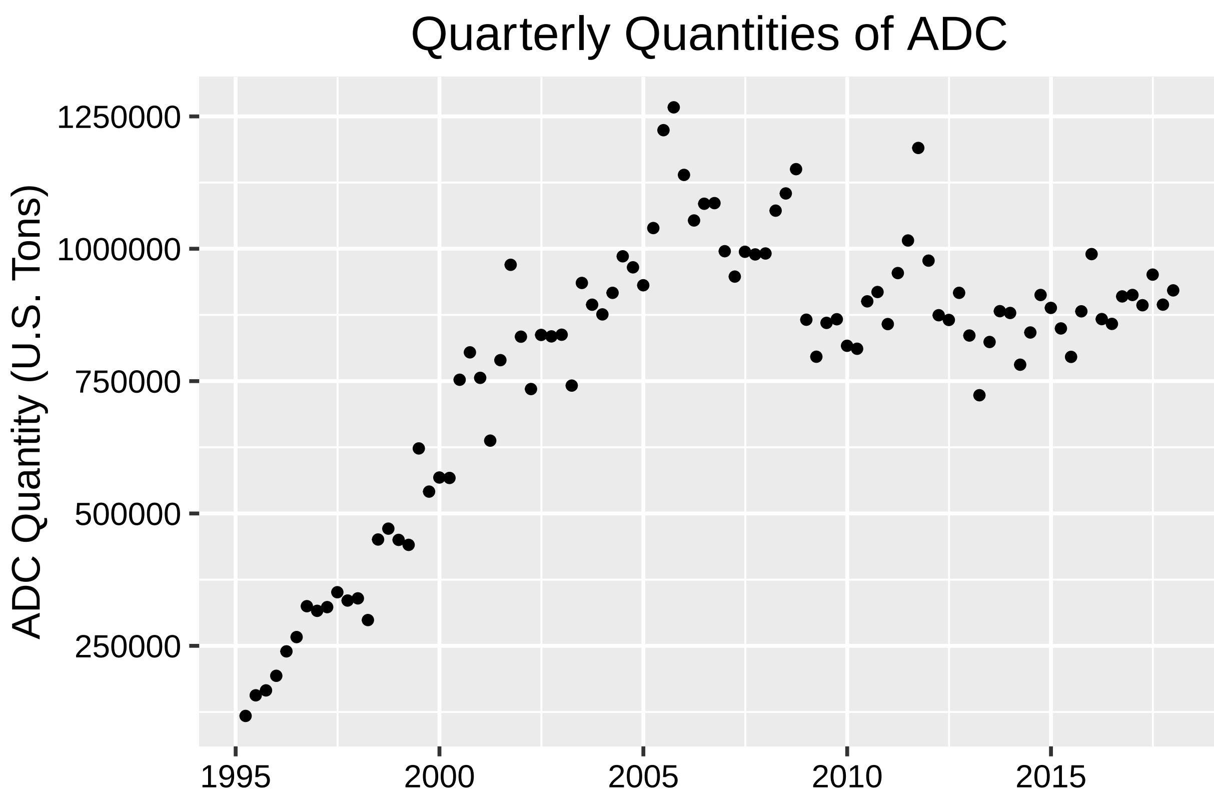
The model uses 1 dependent variable = quarterly quantity, and 1 independent variable = time. The equation was found to be as follows: Quarterly quantity = 73.14*Date - 190264.58. This signifies that for a 1 unit increase in the date, the model predicts there to be a 73.14 ton increase in the ADC quantity.

The adjusted R-squared value was seen to be 0.4433, which means that the model explains 44.33% of the variation. The p-value was found to be 2.694e-13.

```
# assumptions for lm (independent observation, normal distribution,
#equal variances among groups) checked in Test 1. data is not normal,
#but group variances are equal. proceed with lm

# 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', '6-30', '9-30', '12-31'))

# 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
```

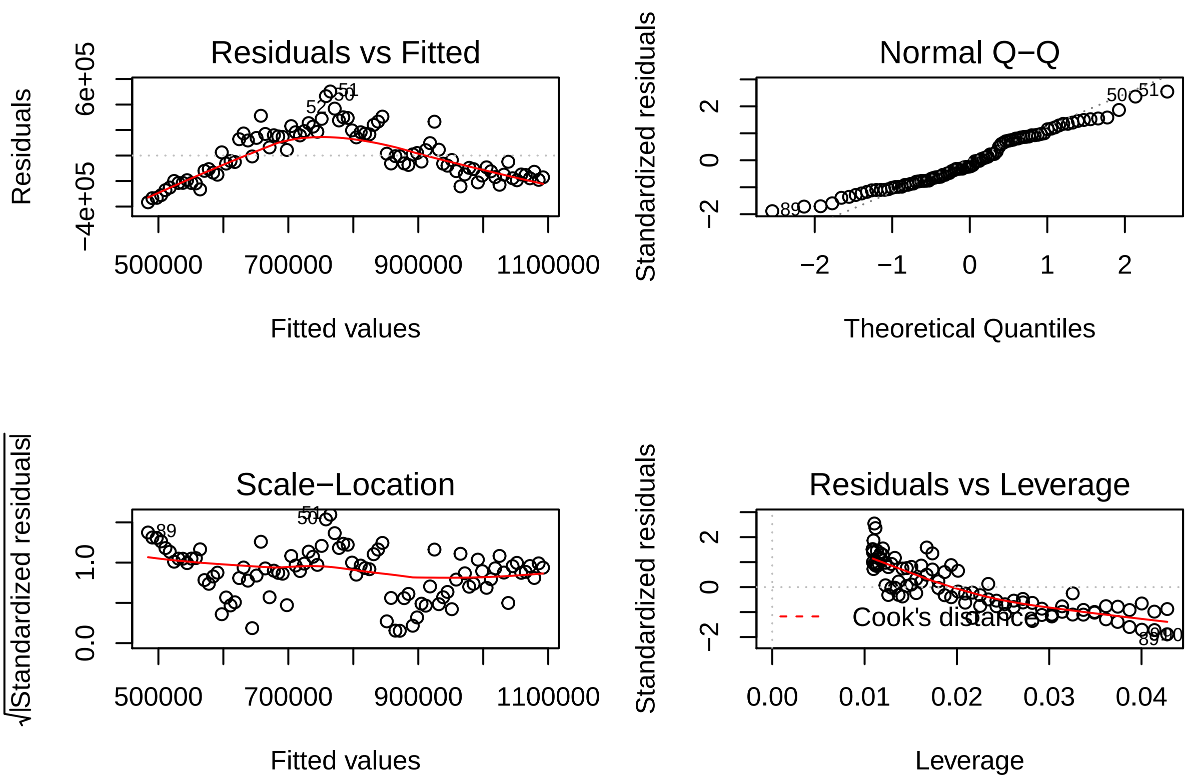
```
##
## Call:
## lm(formula = Total ~ Quarter.End.Date, data = ADC_fulldate)
##
## Coefficients:
##      (Intercept)  Quarter.End.Date
##       -190264.58             73.14
```

```
summary(ADC_date_lm)
```

```
##
## 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
# Adjusted R-squared:  0.4433 (date explains 44.33% of variation in total),
#p-value: 2.694e-13

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



4.2.1 Test 2: Result

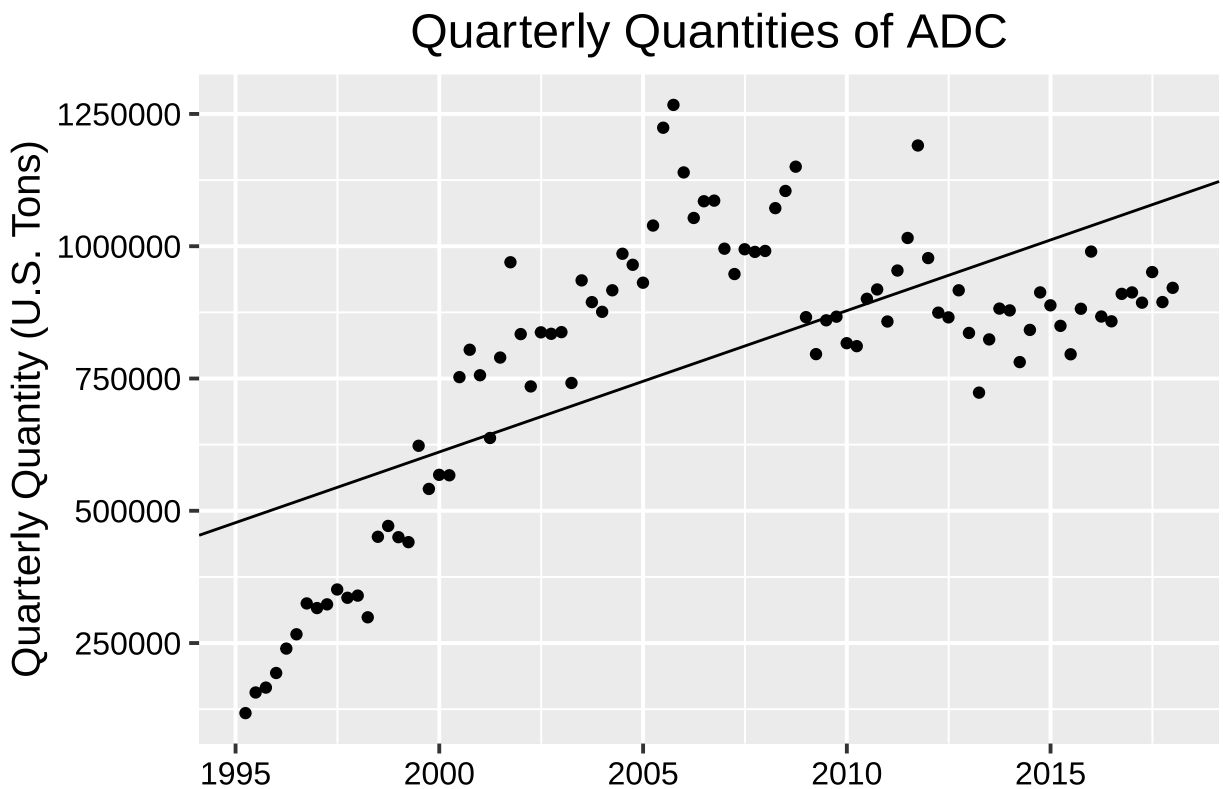


Figure 5: Quarterly quantities of ADC are plotted by time, and fit with the equation y = 73.14*time -190264.58

Figure 5 shows the individual data points plotted over time, overlaid with the model represented by the line. The model does follow the same upward trend that the points show, but visually the model does not appear to be a good fit for the data. The points show a peak around 2006 and level off after that, which is not represented in the model.

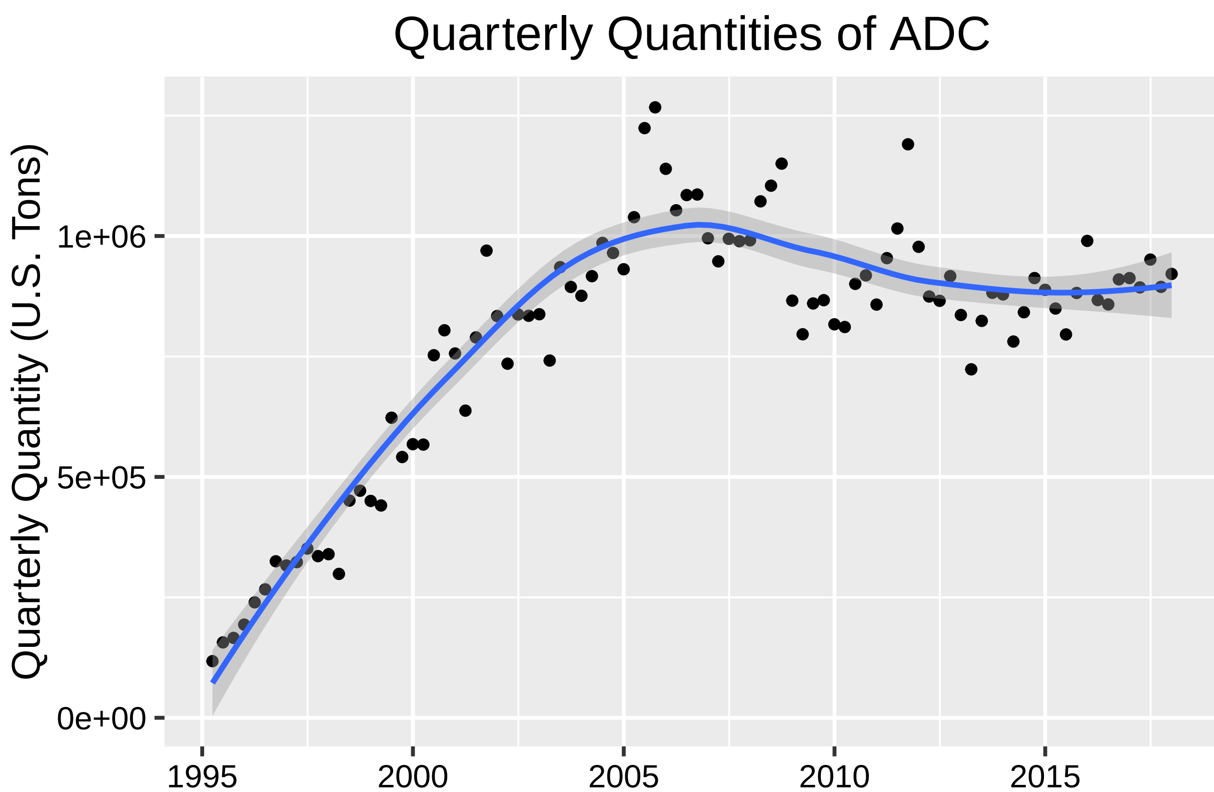


Figure 6: Quarterly quantities of ADC are plotted by time, and fit with a Loess Regression

Figure 6 shows the individual data points plotted over time, overlaid with a Loess regression. The Loess regression appears to be a much better fit to the data points visually. The model is represented by the blue line and the confidence interval of the model is represented by the gray area surrounding the line.

4.3 Test 3: Changepoint in Construction & Demolition - Analysis

Test 3 assessed construction & demolition ADC quantities over time, checking for changepoints in the data. Since this analysis is on type-specific ADC, only data from years 1998-2017 was used. The data was limited to only C&D waste by using the select function.

The determine the type of test that should be used, the following 2 assumptions were evaluated: 1) the observations are independent 2) the groups are normally distributed

Assumption 1 for independent observations cannot be checked from the dataset, and was assumed to be true.

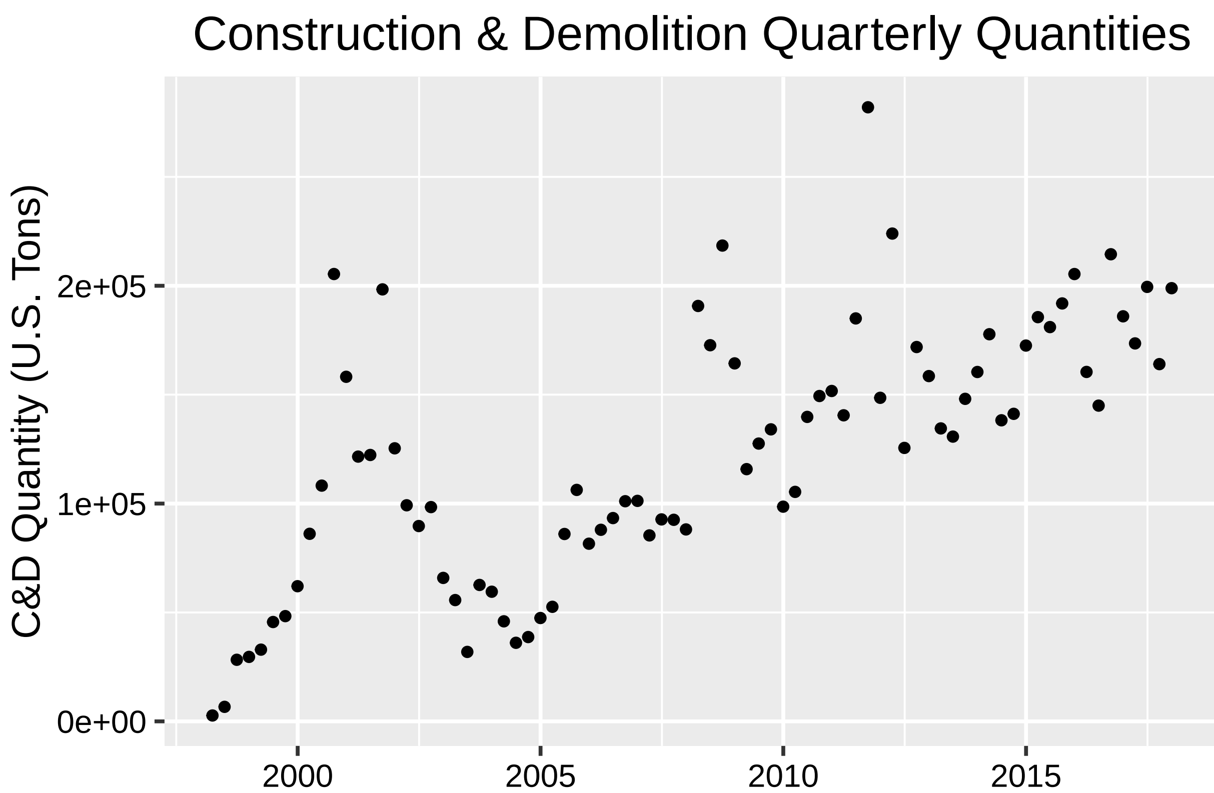
Assumption 2 for normal distribution was checked using the Shapiro Wilks test. The H0 for this test is that of normality, and the p value = 0.4, therefore this data set was deemed to be normally distributed.

A histogram was used to view the distribution, which appears visually to have a normal distribution. A Q-Q plot was also used to check normality, and found that the data matches the 1:1 ratio well.

Although the data was found to be normal, the non-parametric Pettitt test was used to determine a shift in the central tendancy of the time series, because the sample size is not very large. The Pettitt test found p = 3.2e-10, which is < 0.05. The change point was identified at time '40', which corresponds to a time between Q4 2007 & Q1 2008 = ~ 2008-02-14.

A separate Mann-Kendall was run for each section (before and after the change point). In the block of points after time '40', the corresponding p = 0.01308, which indicates that there is likely a second change point in that block. A Pettitt test was run on that chunk and found a second change point corresponding to a time between Q4 2014 & Q1 2015 = ~ 2015-02-14.

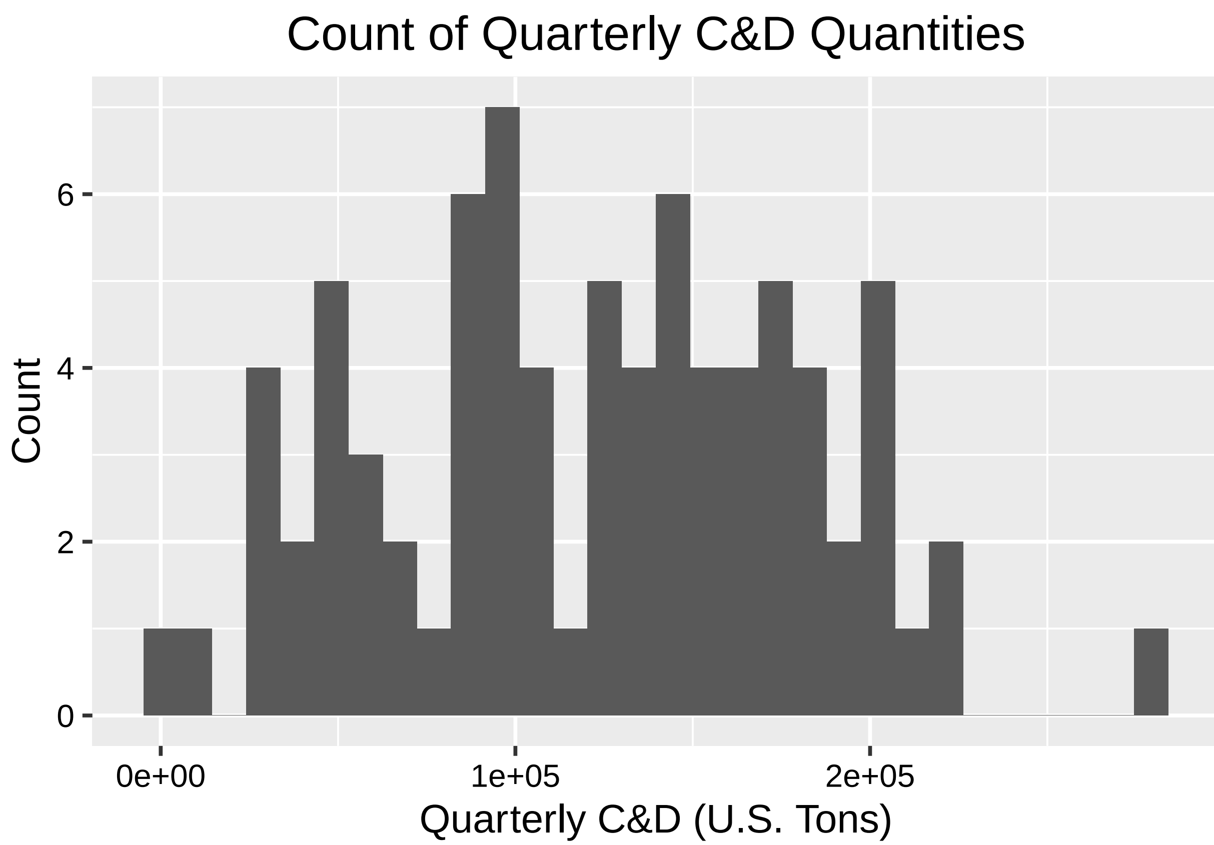
```
# create dataframe with dates
quarters_to_dates$Quarter <- as.integer(quarters_to_dates$Quarter)

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 column as date

# 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")
```



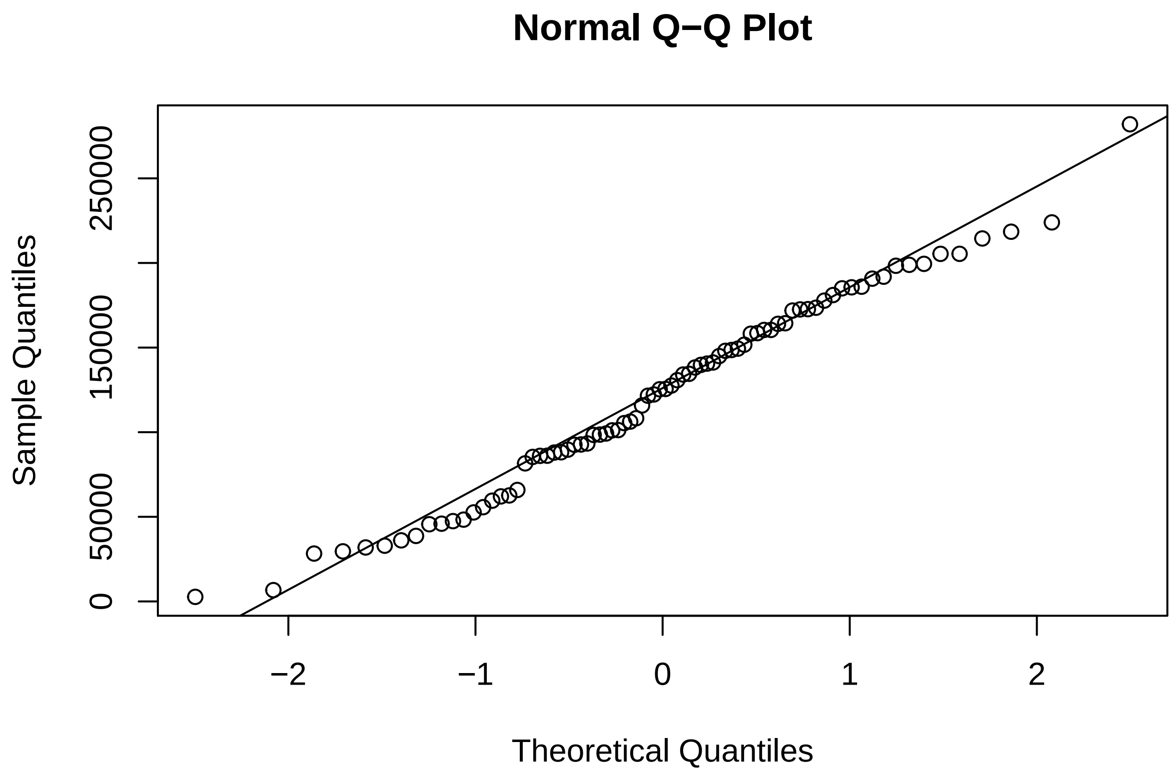

```r
# check normality for C&D waste specifically
shapiro.test(CD_only$Construction.and.Demolition.Waste)
```

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

```r
# p-value = 0.4028, inferring that the data is normal

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.Demolition.Waste) # matches 1:1 ratio pretty well
```




```r
# use Pettitt's test (nonparametric) to determine whether there is a shift
#in the central tendency of the time series.
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
```

```r
# 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
```

```r
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
```

```r
# 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
```

```r
# 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])
```

```
##
##  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
```

```r
# p-value = 0.9453, not likely a 3rd change point
```

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

```
##
##  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
```

```r
# p-value = p-value = 1.261, no 3rd change point

# 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: Result

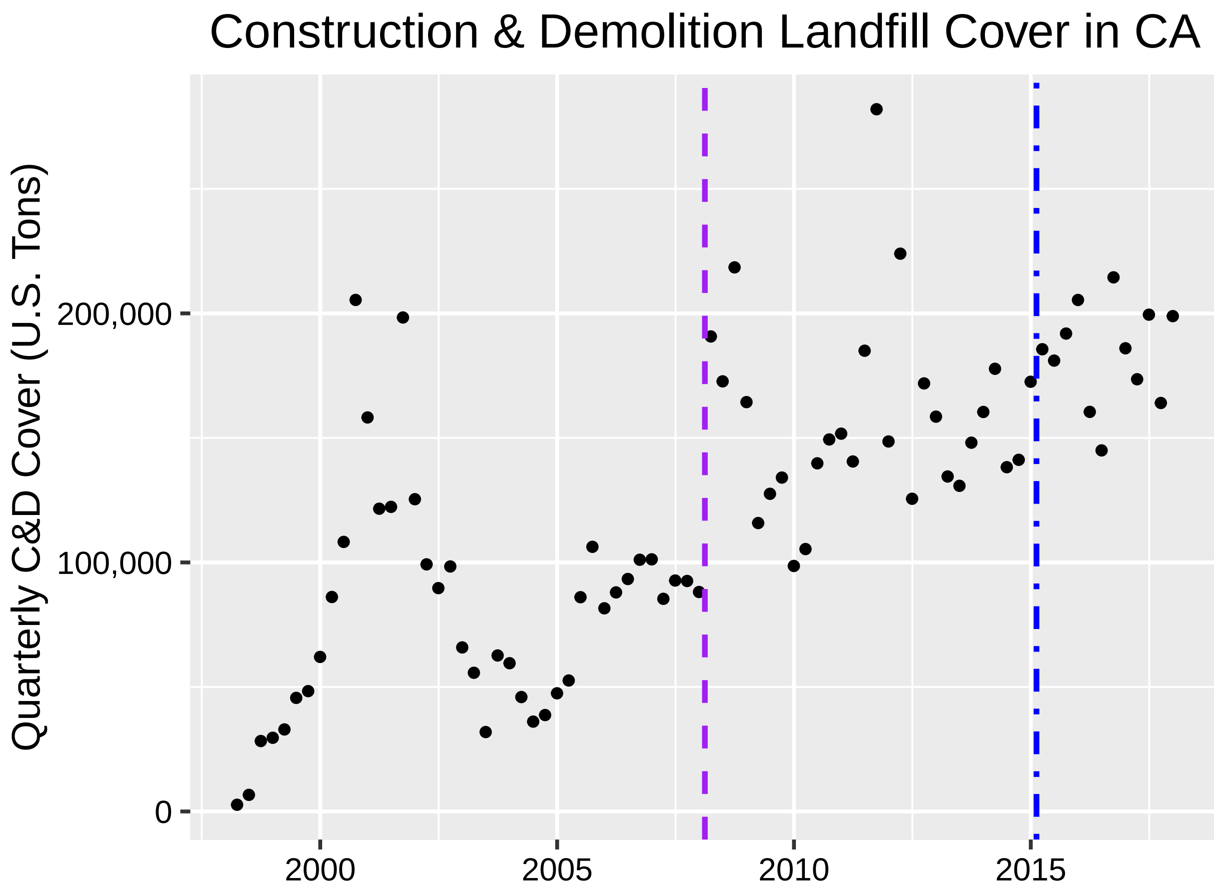


Figure 7: Quarterly quantities of Construction & Demolition ADC are plotted by time, and the 2 changepoints are marked with dotted lines

Figure 7 shows the quarterly C&D quantities plotted over time, with changepoints highlighted at ~ 2008-02-14 (in purple) and ~ 2015-02-14 (blue). There appears to be a changepoint at around 2002, but the statistical test did not detect a changepoint there. It is possible that is due to the sample size. Also, the changepoint detected at ~ 2015-02-14 (blue) does not seem to be represented by the data points. This may also be due to the small sample size, or could be influenced by the high quantity values seen around 2012.

5 Summary and Conclusions

The analysis of the quantity of ADC used in each reporting quarter found that the means of the quarters are not significantly different from one another. This was found quantitatively via the dunn post-hoc test, but can also be seen visually from a violin plot. This lack of a quarterly trend infers that the supply of ADC (or of earthen material) is relatively constant across the year. This could be a nationwide trend, or a special effect from California's temperate climate allowing for relatively similar weather (and thus growing season) year-round.

When attempting to model quarterly ADC quantities as a function of time the linear model was a moderate fit, explaining ~44% of the variation. Visually though, the model equation (Quarterly Quantity = 73.14*Date - 190264.58) did not appear to be a good fit for the data points. The Loess regression was seen to provide a much better fit visually. For future studies, models beyond the y-mx+b equation would likely be better predictors of quantity of ADC over time.

For construction & demolition ADC specifically, there were 2 changepoints in the quantities over time - one around 2008-02-14, and another around 2015-02-14. Literature research did not show an association of either of these dates with a major change in regulation surrounding landfill diversion standards, but there may be other policy factors influencing these findings. The trends in C&D waste would be useful information to share with landfill operators to help them predict influxes in the production of hydrogen sulfide.

From both the data exploration as well as the statistical analysis, it is clear that various trends exist in the use of ADC across the state of California. Analyses from this data can be used to influence policy makers on sustainability standards, and operate landfills more effectively.