Analysis_DielArrival

Anna Steel

October 21, 2016

Read in non-rediscretized data (just filtered, prior to splitting into bursts) and add release metadata

```
dat <- readRDS("Maestros/AllFish_FiltSec3Hold.RData")

fishrel = read.csv("C:/Users/Anna/Documents/GitHub/Fremont16/Maestros/TagReleaseList.csv",
                   colClasses=c("RelTime"="character"))
  fishrel$datetime = as.POSIXct(paste(fishrel$RelDate, fishrel$RelTime), format="%Y-%m-%d %H%M", tz="
  names(fishrel)[3] <- "id"
  names(fishrel)[ncol(fishrel)] <- "datetime.Rel"

 dat = merge(dat, fishrel, all.x=T)
  dat$RelHr = as.POSIXlt(dat$datetime.Rel)$hour
```

Store sunset and sunrise times:

- Referenced from: http://aa.usno.navy.mil (mean for range of all detections, rel 1-5: 2/22 - 3/19/2016)

```
sunrise = 06.49
sunset = 18.07
```

Store distance between Tisdale Weir release site and top of receiver array

- Calculated from Google Earth positions, referencing CFTC receivers at known locations / rkm
- The rkm are calculated with the golden gate at rkm 0, and Chipps Island at km 69.5

```
travdist_km = 55.1
```

Run several more data cleaning and organizing steps

Extract first and last detections for each fish

Calculate hour of day (decimal hours) for time when fish were released & when fish arrived at array; calculate transit time (a.k.a. 'delay') between

- Add code for night or day, using sunrise/sunset times incorporated above, to both release and arrival.
- Also calculate passage time - may be useful later

Calculate mean and median transit times

```
## [1] "mean = 37.9104605274909"
```

```
## [1] "sd = 22.8904213895797"
```

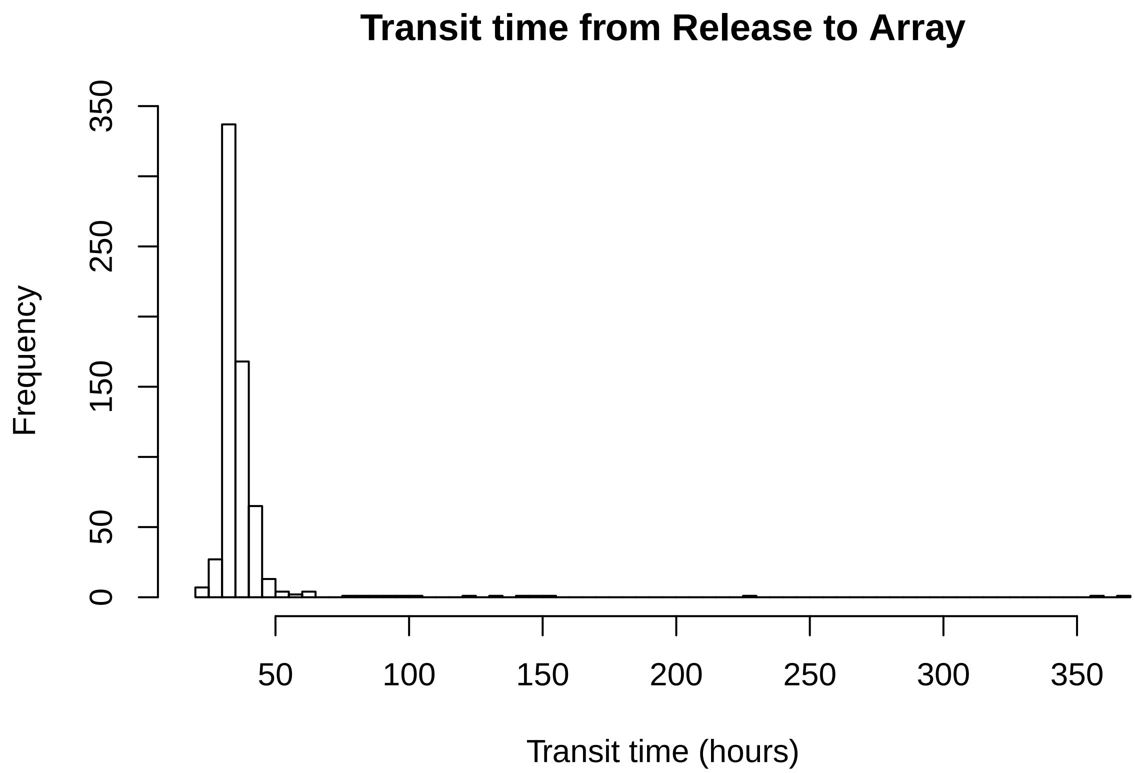
```
## [1] "median = 34.356684213082"
```

```
##   RelEv  mean_hr     sd_hr median_hr
## 1     1 44.45239 37.514802  35.42802
## 2     2 36.46106  5.662311  36.48692
## 3     3 32.97934  8.174744  31.98577
## 4     4 32.08457  1.799459  31.77429
## 5     5 32.20419  6.365311  31.53557
```

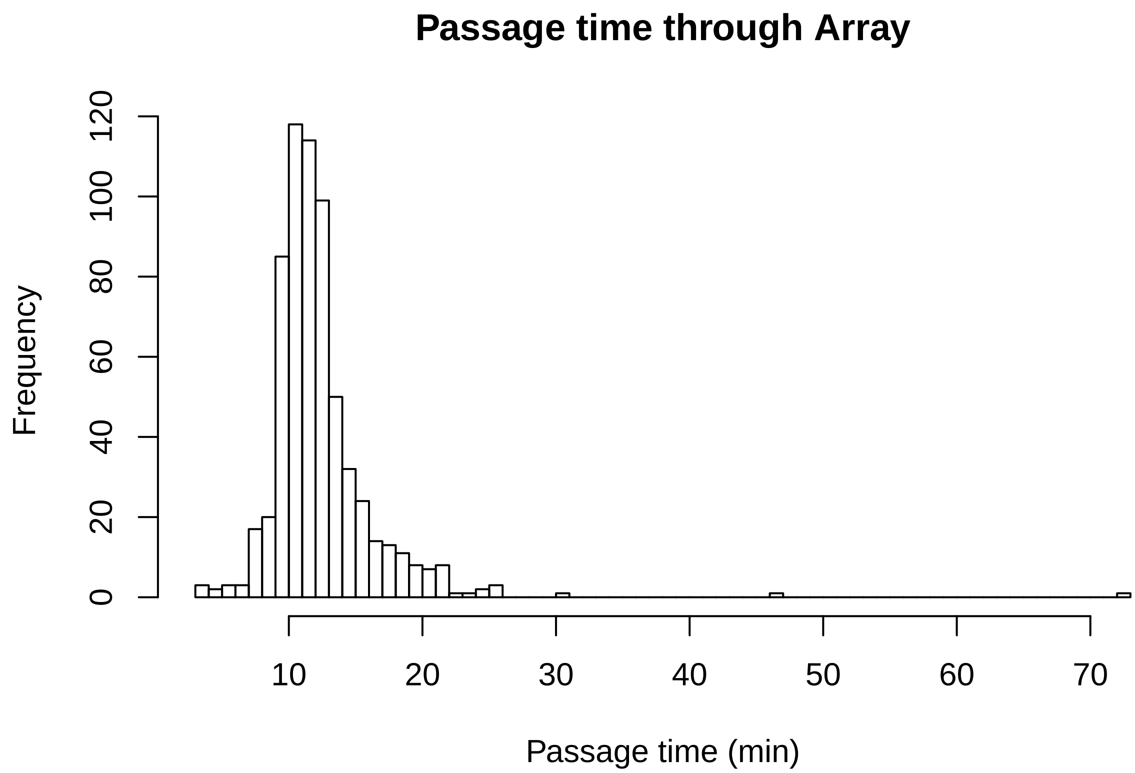
```
##    RelEv Rel.hrDayfac  mean_hr     sd_hr median_hr
## 1      1            6 43.57106 46.441281  34.98257
## 2      1           11 43.44061 46.753894  34.37964
## 3      1           17 49.19190 34.780700  40.32683
## 4      1           23 41.87227 12.894312  37.07309
## 5      2            3 41.15825  5.917072  40.03887
## 6      2            9 38.44610  2.976854  37.70178
## 7      2           15 35.63652  5.233195  33.69668
## 8      2           21 30.93178  1.061992  30.73722
## 9      3            0 32.18494  1.693167  32.04584
## 10     3            4 34.67237 13.879419  32.10779
## 11     3            8 32.05683  1.461632  31.81823
## 12     4            2 31.69791  1.345745  31.31554
## 13     4           10 32.25642  2.100366  31.87870
## 14     4           18 32.25643  1.896995  32.47462
## 15     5            2 25.63799  3.646073  25.15544
## 16     5           10 31.59138  1.188160  31.24606
## 17     5           18 39.88828  1.092754  39.59271
```

Plots to visualize initial transit time and passage time

```
hist(fl.df2$Delay.hr,
     main="Transit time from Release to Array",
     xlab="Transit time (hours)", ylab="Frequency", breaks=50)
```

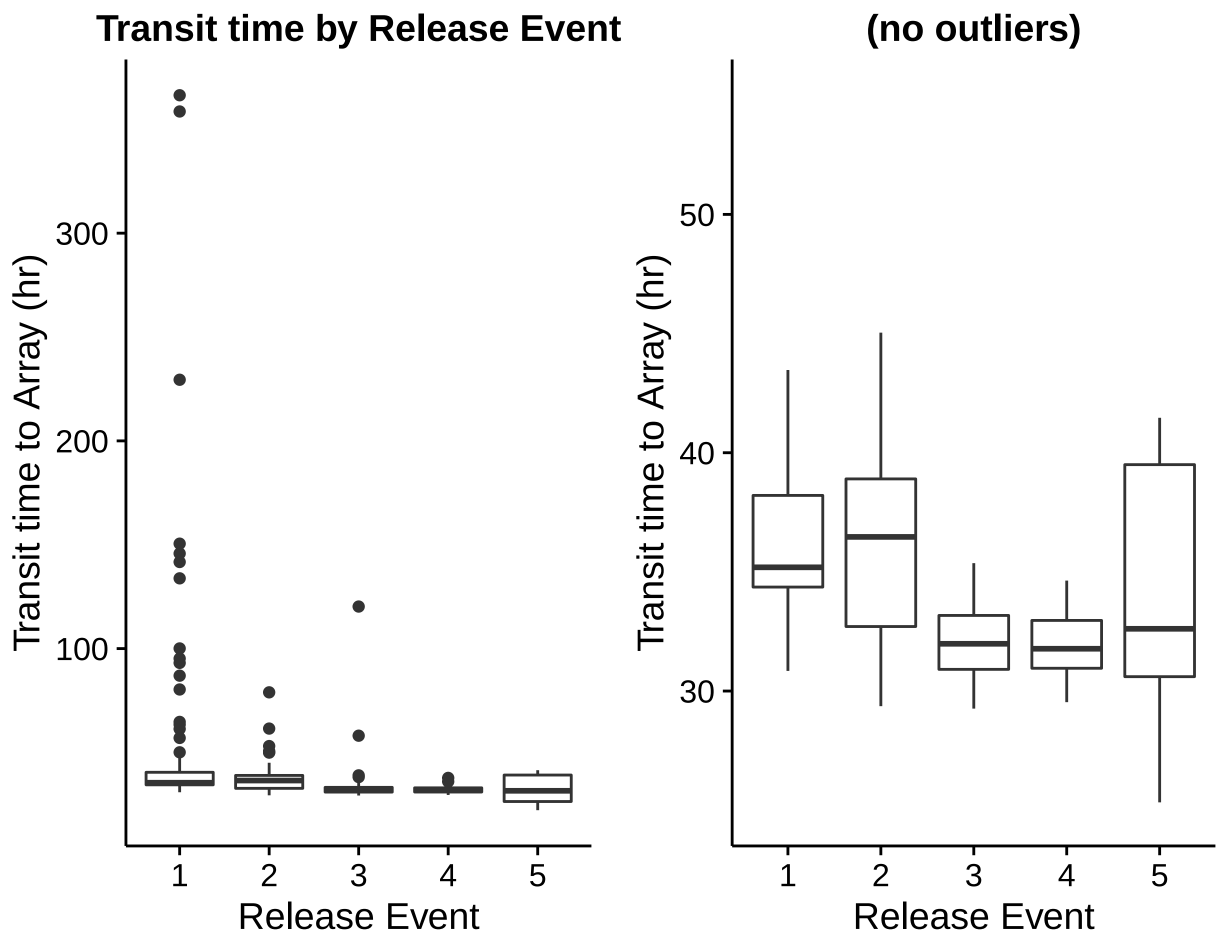


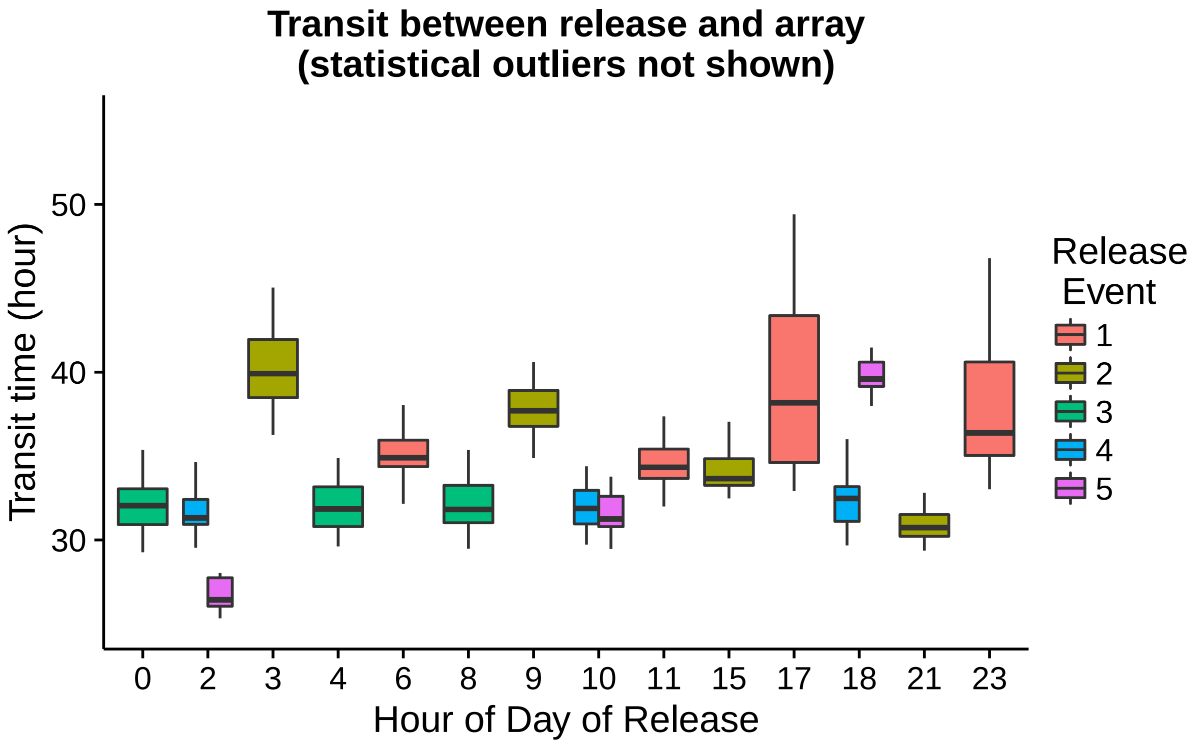

```r
hist(fl.df2$passtime.min,
     main="Passage time through Array",
     xlab="Passage time (min)",ylab="Frequency", breaks=50)
```

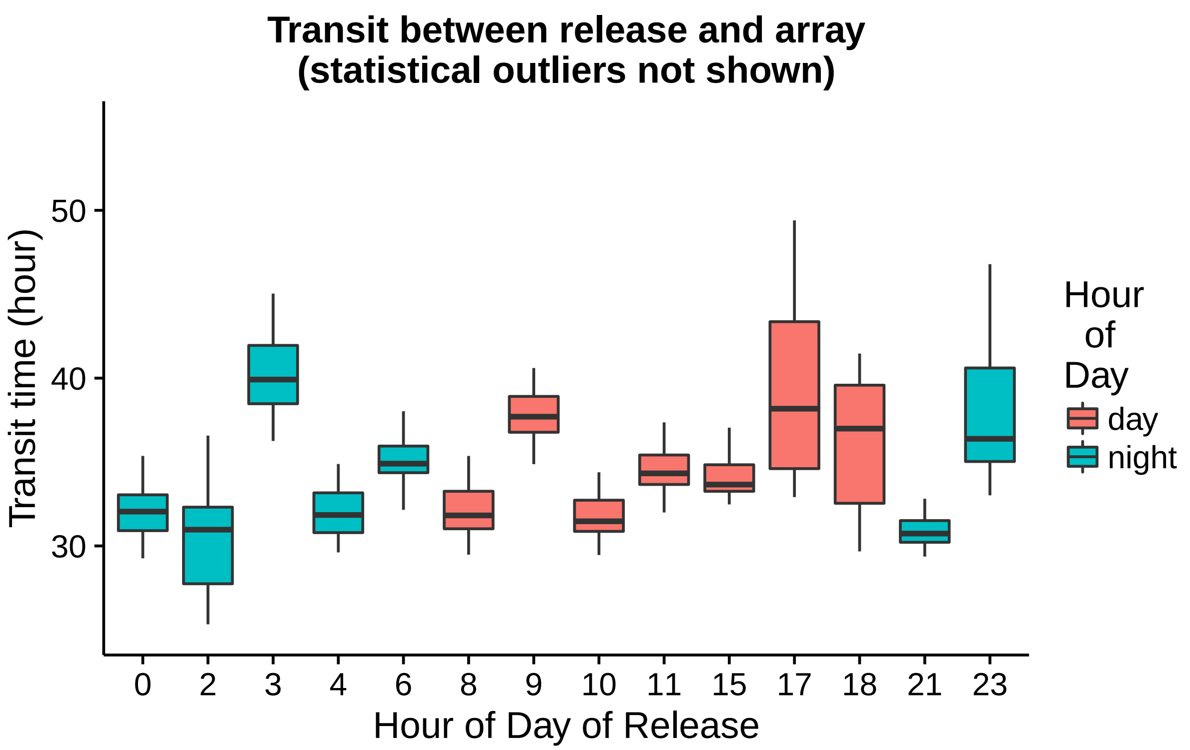


Differences by Release Event or Release Hour

- note: it would be nice to annotate boxes with respective sample sizes


Transit time by Release Event
(no outliers)
Transit time to Array (hr)
300
200
100
1
2
3
4
5
Release Event
50
40
30
Transit time to Array (hr)
1
2
3
4
5
Release Event





- note: consider creating this second set of boxplots in conjunction with a hydrograph to illustrate relationship between transit time and stage; from these boxplots that seems like it might be the primary driver

Mean and Median of arrival times, overall and by release time groups

- requires circular statistics; use packages psych (mean/sd) and circular (median)

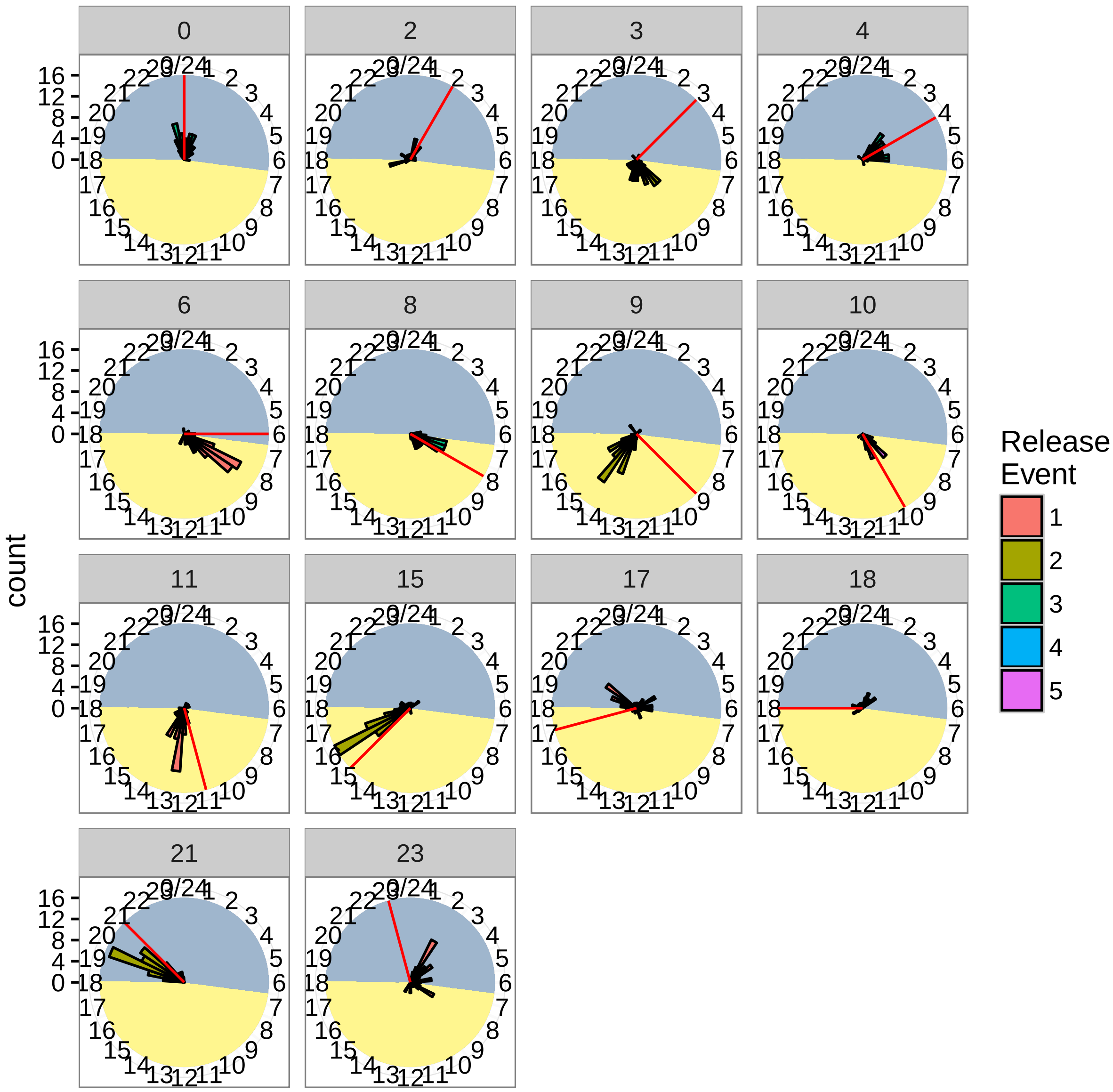
```
## [1] "mean = 10.7477423488094"
```

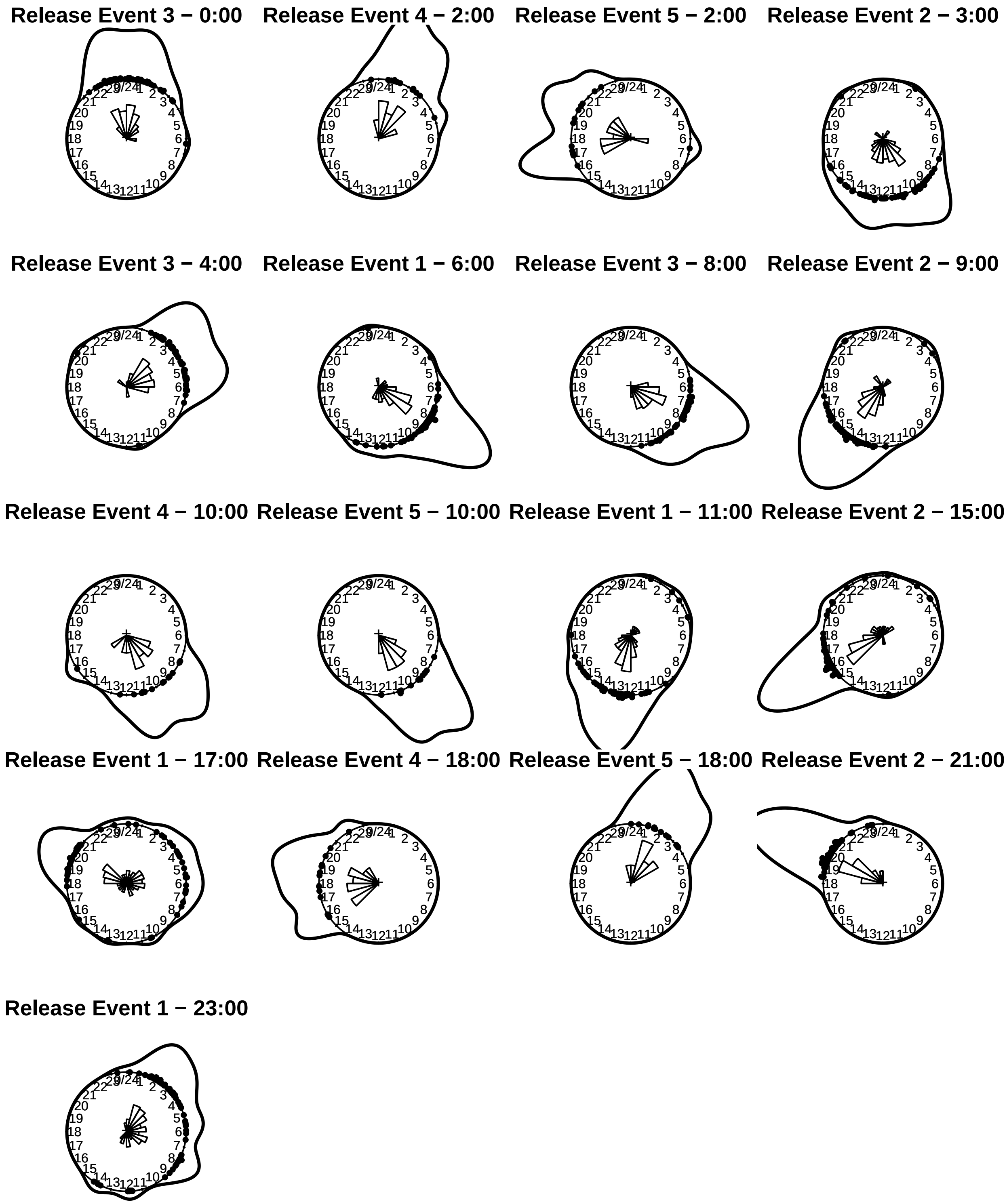
```
## [1] "sd = 2.15151083557221"
```

```
## [1] "median = 10.3"
```

```
##    Rel.hrDayfac median.F.hr  mean.F.hr   sd.F.hr
## 1             0     0.20000  0.2670601 0.4330199
## 2             2    22.75000 22.4047524 1.1392728
## 3             3    10.95000 11.2855469 0.6970469
## 4             4     3.90000  4.0722226 0.5285302
## 5             6     8.40000  8.6452943 0.5289553
## 6             8     7.75000  8.0408195 0.3792161
## 7             9    14.50000 14.7191705 0.6036528
## 8            10     9.50000  9.8554812 0.4329048
## 9            11    12.72000 13.0603746 0.6941681
## 10           15    16.30000 16.7919601 0.6984048
## 11           17    20.63333 23.2063704 1.5831653
## 12           18    23.00000 22.2926143 1.1737111
## 13           21    19.80000 20.0129941 0.2794529
## 14           23     3.85000  4.5392087 0.9462162
```

Circular plots: same data, two visualizations





- the blobs around the circle are kernel density lines, but the smoothing parameter is simply the default; if these graphics are going to be used for anything other than general exploration of the data I should revisit the smoothing parameter selection process.

Preliminary Exploration of circular statistics for diel questions

Initial look at anova and diagnostics for the lm to assess relationship between release event (1-5) and release time on transit time

Can we evaluate fish released at different times all together, or do we need to seperate them out?

Indicates that only release event is important, but **doesn't meet assumptions of normality**. Also, doesn't include the reality that the release time is a circular value.

```
summary(aov(fl.df2$Delay.hr~factor(fl.df2$RelEv) * fl.df2$Rel.hrDay))
```

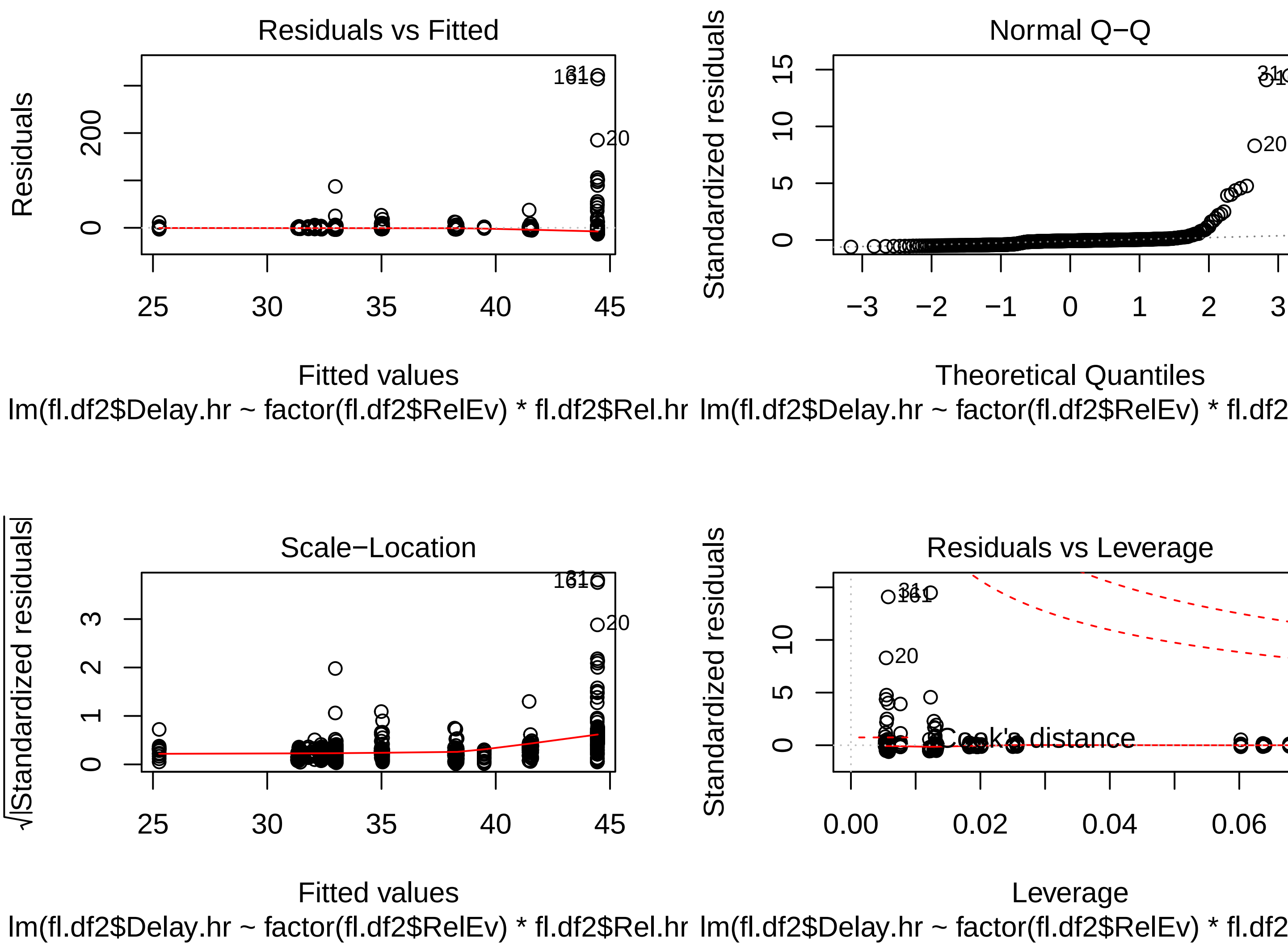
```
##                                        Df Sum Sq Mean Sq F value   Pr(>F)
## factor(fl.df2$RelEv)                    4  15507    3877   7.755 4.19e-06
## fl.df2$Rel.hrDay                        1    611     611   1.221     0.27
## factor(fl.df2$RelEv):fl.df2$Rel.hrDay   4   3783     946   1.892     0.11
## Residuals                             631 315440     500
##
## factor(fl.df2$RelEv)                  ***
## fl.df2$Rel.hrDay
## factor(fl.df2$RelEv):fl.df2$Rel.hrDay
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(fl.df2$Delay.hr~factor(fl.df2$RelEv) * fl.df2$Rel.hrDay))
```

```
##
## Call:
## lm(formula = fl.df2$Delay.hr ~ factor(fl.df2$RelEv) * fl.df2$Rel.hrDay)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -13.61  -4.57  -1.44   0.12 321.94
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             44.470940   3.552814  12.517
## factor(fl.df2$RelEv)2                   -1.208679   4.752330  -0.254
## factor(fl.df2$RelEv)3                  -11.446789   4.726757  -2.422
## factor(fl.df2$RelEv)4                  -12.737480   7.616629  -1.672
## factor(fl.df2$RelEv)5                  -20.977761   7.275299  -2.883
## fl.df2$Rel.hrDay                        -0.001315   0.227584  -0.006
## factor(fl.df2$RelEv)2:fl.df2$Rel.hrDay  -0.561417   0.322610  -1.740
## factor(fl.df2$RelEv)3:fl.df2$Rel.hrDay  -0.009907   0.649650  -0.015
## factor(fl.df2$RelEv)4:fl.df2$Rel.hrDay   0.035720   0.603785   0.059
## factor(fl.df2$RelEv)5:fl.df2$Rel.hrDay   0.890193   0.584260   1.524
##                                        Pr(>|t|)
## (Intercept)                             < 2e-16 ***
## factor(fl.df2$RelEv)2                   0.79932
## factor(fl.df2$RelEv)3                   0.01573 *
```

```
## factor(fl.df2$RelEv)4                          0.09496 .
## factor(fl.df2$RelEv)5                          0.00407 **
## fl.df2$Rel.hrDay                               0.99539
## factor(fl.df2$RelEv)2:fl.df2$Rel.hrDay  0.08231 .
## factor(fl.df2$RelEv)3:fl.df2$Rel.hrDay  0.98784
## factor(fl.df2$RelEv)4:fl.df2$Rel.hrDay  0.95284
## factor(fl.df2$RelEv)5:fl.df2$Rel.hrDay  0.12810
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.36 on 631 degrees of freedom
## Multiple R-squared:  0.05935,    Adjusted R-squared:  0.04593
## F-statistic: 4.423 on 9 and 631 DF,  p-value: 1.243e-05
```

```
mfrow=c(2,2); plot(lm(fl.df2$Delay.hr~factor(fl.df2$RelEv) * fl.df2$Rel.hrDay))
```



The following are derived from the test statistics I ran for 2015 to compare runs

- Much/all of the following was coded with guidance from the text book "Circular Statitics in R' by Arthur Pewsey et al (2013)

To select an appropriate statistical distrubition for the circular data we want to know if the data are symetrical (we know they are not uniform from looking at the plots, so won't bother to test this statistically, for now). If we do not reject symmetry, we may use the Jones-Pewsey or vonMises distributions, but if we do reject symmetry we may need to use the more flexible Batschelet distribution.

Test for 'reflective symmetry'

Depending upon the nature of the data, we can or can't use some of the following circular statistical tests We can use he test proposed by Pewsey (2002) which is suitable for sapmle sizes of 50 or more (ours are n=51 - 56 in each release hour)

```
##    Relhrfac   teststat        pval
## 1         0 1.32324811 0.185752882
## 2         2 0.25281056 0.800414604
## 3         3 0.80442009 0.421154403
## 4         4 0.05209211 0.958455302
## 5         6 0.34426924 0.730643811
## 6         8 1.68316207 0.092343721
## 7         9 0.97181120 0.331144484
## 8        10 1.11375442 0.265384554
## 9        11 1.15390020 0.248541089
## 10       15 2.38374045 0.017137685
## 11       17 1.08282359 0.278886734
## 12       18 0.45383879 0.649944863
## 13       21 2.25633658 0.024049560
## 14       23 3.16859376 0.001531783
```

- NOTE: this uses template=clock24 and rotation=clock which may pose a problem; The functions may be expecting radians measured *counter-clockwise* from zero (in mathematic terms, so zero = *positive X-axis*). Here I use radians measured *clockwise* from the *top of the unit circle*. Before moving along or using these values, clarify this.
- Aside from that concern, the **results are mixed** -> releases at 15:00 (rel2), 21:00 (rel2) and 23:00 (re11) are not statistically symmetrical, but the release at 17:00 (rel1) is, dispite not resembling a normal distribution but rather being bimodal. Interesting. Not sure if any of this will be used in a report, so I'm not pursuing it at this moment.

Is the delay in arrival time related to release time?

I tried to use the guidance in Pewsey 2013 textbook to fit a cosine regression model, but the data on delay time are too skewed for it to fit the assumptions. Additionally, I'm not sure it's clear what the model will tell you?

Perhaps this indicates that there is NOT a significant effect of release time on travel time - ie: there isn't a strong or clear diel effect.

Regardless, I'm also not sure if time sunk into this exercise is valuable, so it will be put on the back burner for now.

Basic cosine model:

```r
# calculate a circular correlation as first pass at this relationship
circadian.linear.cor(fl.df2$Delay.hr, fl.df2$Rel.hrDay)
```

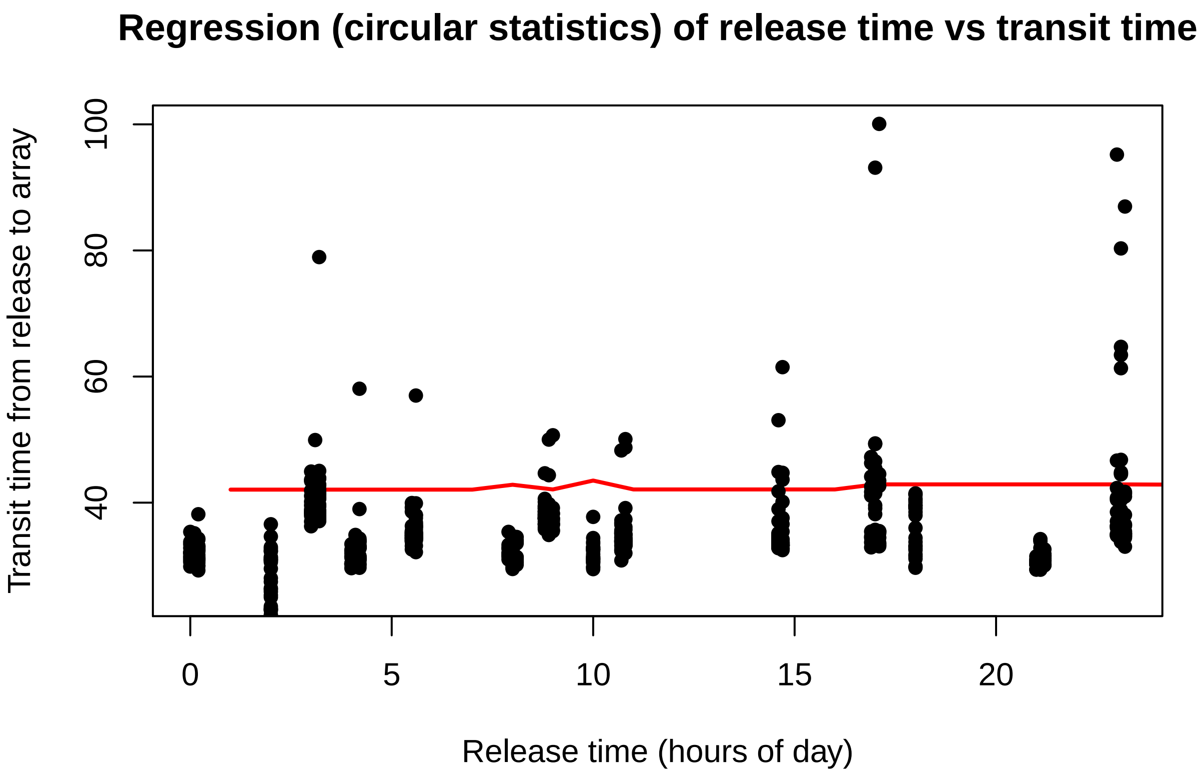
```
## [1] 0.06010921
```

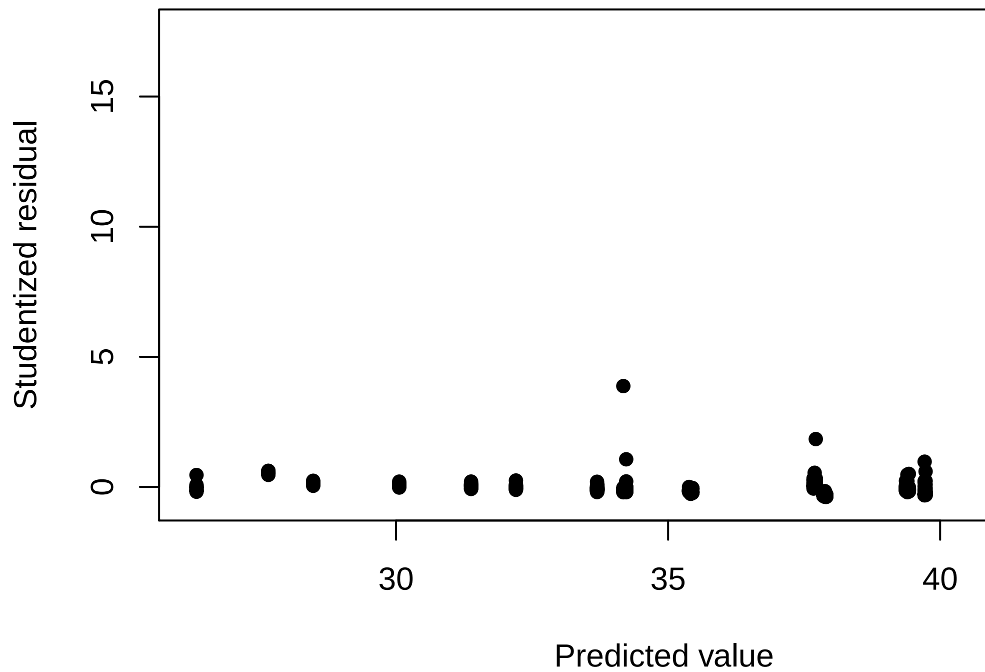
```r
# Next step: regression of a linear response on a circular predictor
# basic cosine model: x = a + b1*cos(2pi/24*Rel.hr*Day) + b2*sin(2pi/12*Rel.hrDay) + e
omega = 2*pi/24
cosvar = cos(omega*fl.df2$Rel.hrDay)
sinvar = sin(omega*fl.df2$Rel.hrDay)

delaymod = lm(fl.df2$Delay.hr ~ fl.df2$RelEv + cosvar + sinvar)
summary(delaymod)
```

```
##
## Call:
## lm(formula = fl.df2$Delay.hr ~ fl.df2$RelEv + cosvar + sinvar)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -12.66  -6.55  -2.88   0.38 324.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   46.1187     1.9076  24.176  < 2e-16 ***
## fl.df2$RelEv  -3.7263     0.7819  -4.765 2.34e-06 ***
## cosvar        -1.2385     1.2695  -0.976    0.330
## sinvar        -0.1653     1.2894  -0.128    0.898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.51 on 637 degrees of freedom
## Multiple R-squared:  0.03763,    Adjusted R-squared:  0.03309
## F-statistic: 8.302 on 3 and 637 DF,  p-value: 2.013e-05
```

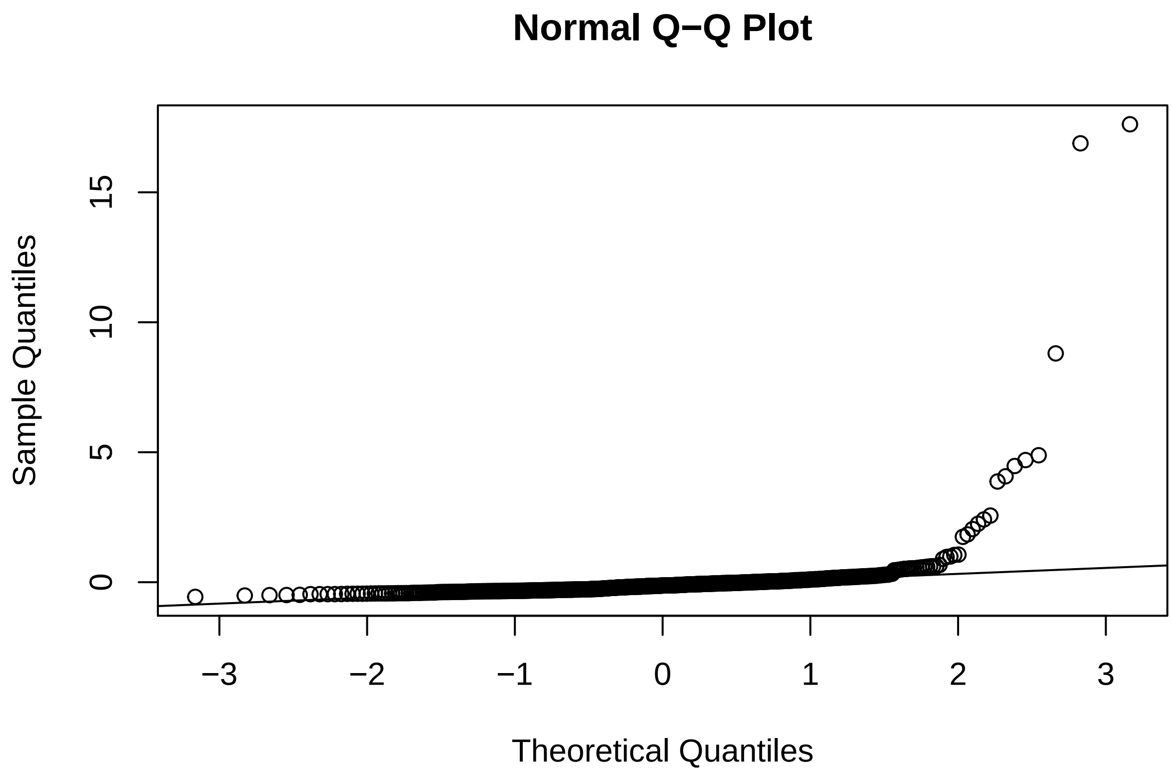
```r
plot(fl.df2$Rel.hrDay, fl.df2$Delay.hr,
     xlab="Release time (hours of day)",
     ylab="Transit time from release to array",
     main="Regression (circular statistics) of release time vs transit time",
     pch=16, ylim=c(25,100),
     lines(predict(delaymod), lwd=2, col="red") )
```


Regression (circular statistics) of release time vs transit time
Transit time from release to array
100
80
60
40
0
5
10
15
20
Release time (hours of day)



Diagnostics of basic cosine model:

```
##
##  Shapiro-Wilk normality test
##
## data:  delayresid
## W = 0.21922, p-value < 2.2e-16
```




```
##
##  Bartlett test of homogeneity of variances
##
## data:  delayresid and fl.df2$Rel.hrDay
## Bartlett's K-squared = 1913.7, df = 32, p-value < 2.2e-16
```

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  delayresid and fl.df2$Rel.hrDay
## Fligner-Killeen:med chi-squared = 167.62, df = 32, p-value <
## 2.2e-16
```

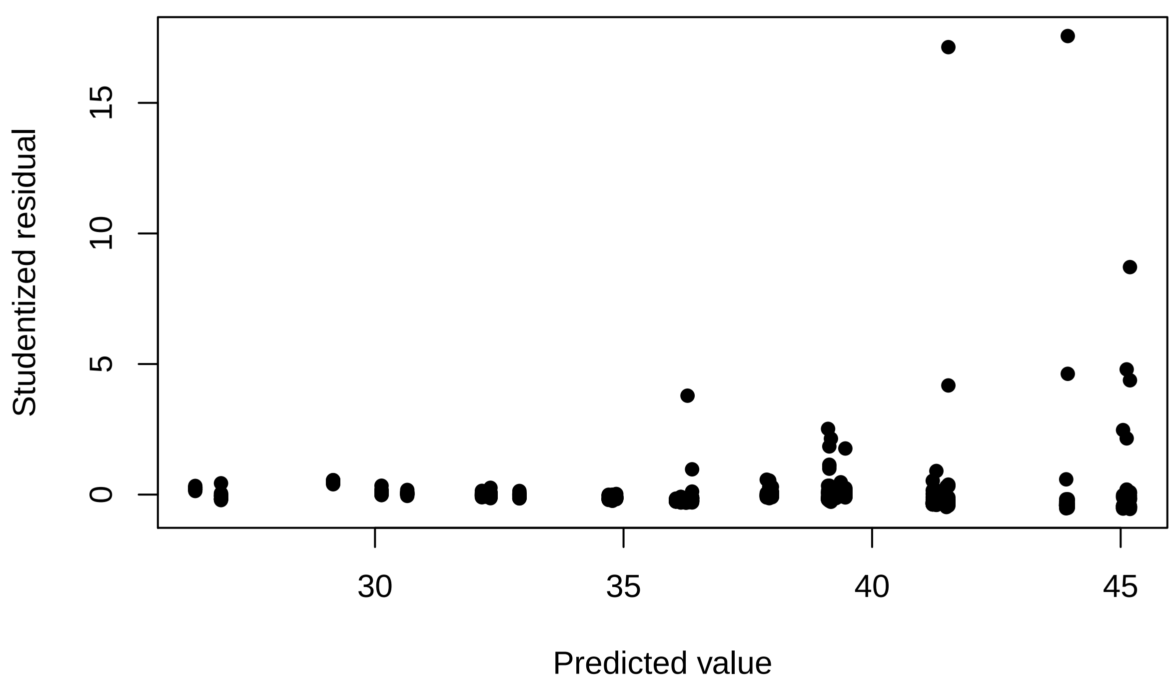
Doesn't meet assumptions, due to outliers with long delay times. But is it close enough?

Extended cosine model (additional sin & cos parameters):

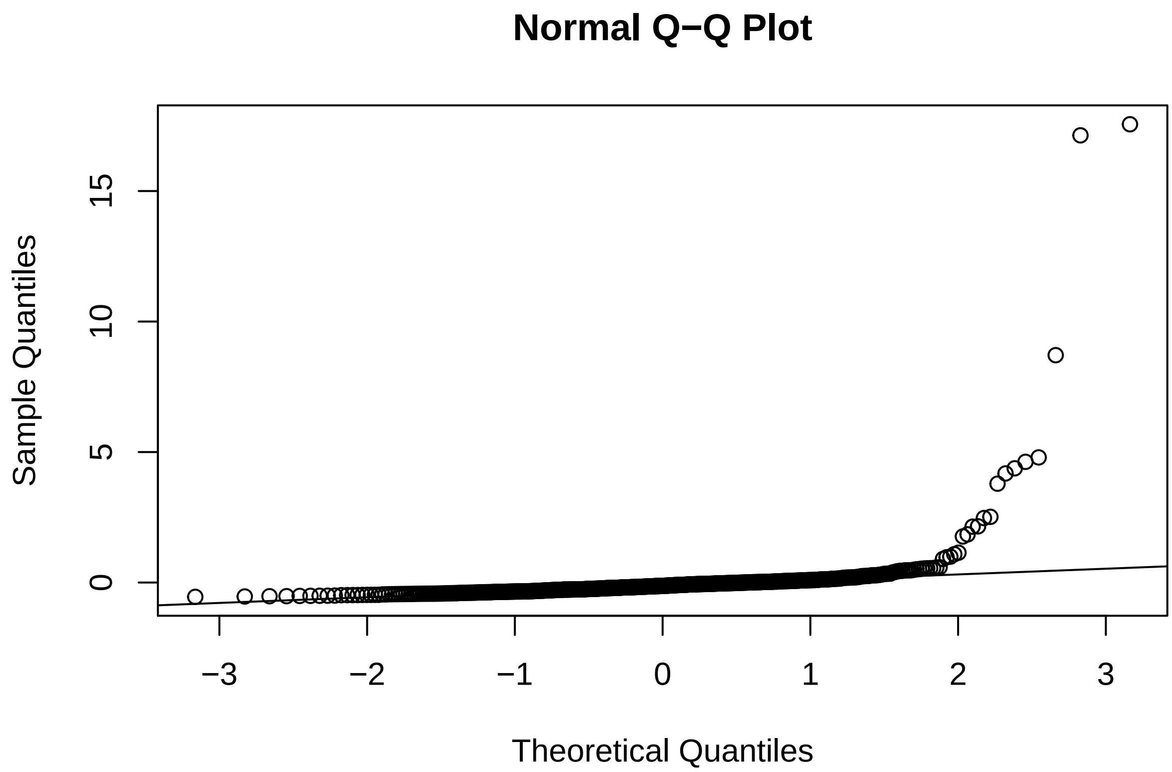
```
##
## Call:
## lm(formula = fl.df2$Delay.hr ~ fl.df2$RelEv + cosvar + sinvar +
##     cos2var + sin2var)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -12.28  -6.12  -2.86   0.47 322.47
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   46.2174     1.9082  24.221  < 2e-16 ***
## fl.df2$RelEv  -3.7471     0.7814  -4.795 2.02e-06 ***
## cosvar        -1.3556     1.3311  -1.018    0.309
## sinvar        -0.2069     1.2918  -0.160    0.873
## cos2var       -1.4697     1.4287  -1.029    0.304
## sin2var        1.6549     1.2629   1.310    0.191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.48 on 635 degrees of freedom
## Multiple R-squared:  0.04345,    Adjusted R-squared:  0.03592
## F-statistic: 5.769 on 5 and 635 DF,  p-value: 3.219e-05
```

Diagnostics of extended cosine model:




```
##
##  Shapiro-Wilk normality test
##
## data:  delay2resid
## W = 0.22178, p-value < 2.2e-16
```




```
##
##  Bartlett test of homogeneity of variances
##
## data:  delay2resid and fl.df2$Rel.hrDay
## Bartlett's K-squared = 1915.1, df = 32, p-value < 2.2e-16
```

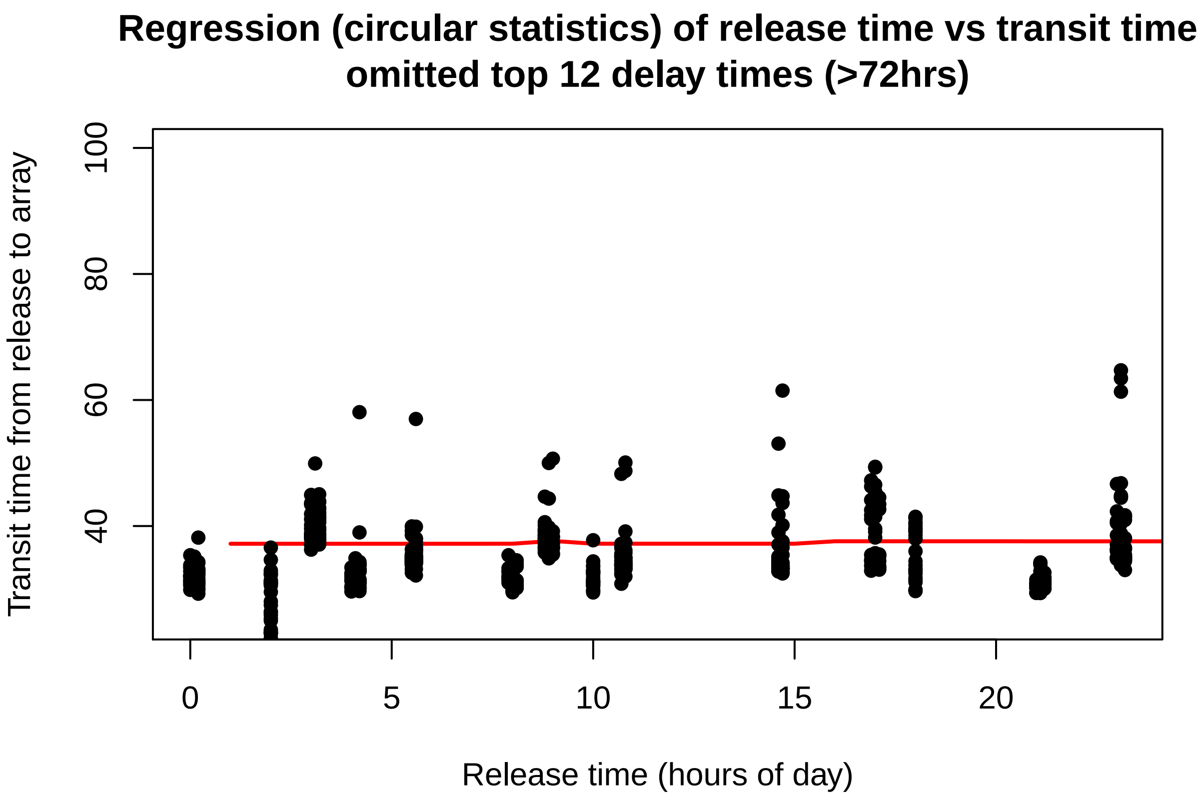
```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  delay2resid and fl.df2$Rel.hrDay
## Fligner-Killeen:med chi-squared = 167.71, df = 32, p-value <
## 2.2e-16
```

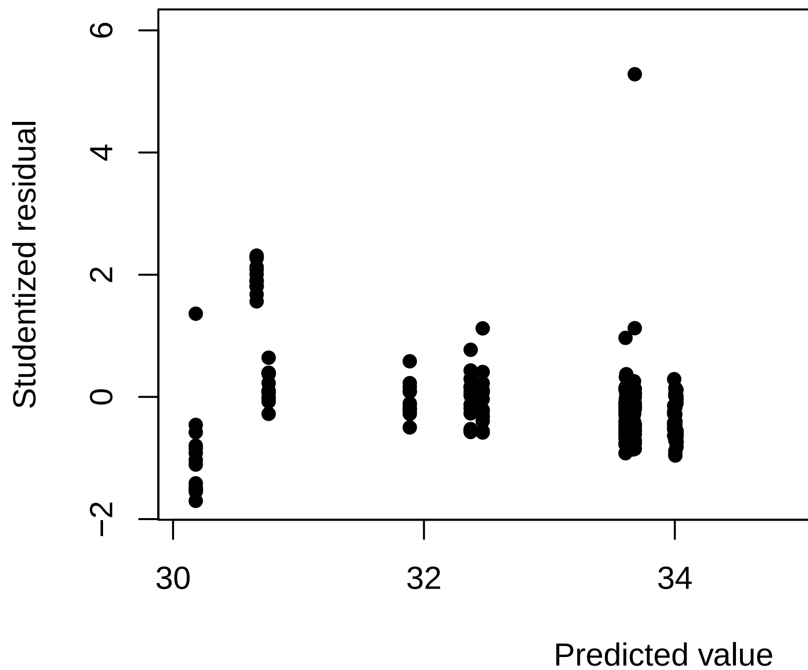
Still doesn't meet assumptions well.

Extended cosine model without outliers

```
##
## Call:
## lm(formula = fl.df4$Delay.hr ~ fl.df4$RelEv + cosvar + sinvar)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.979 -2.890 -1.311  1.816 27.645
```

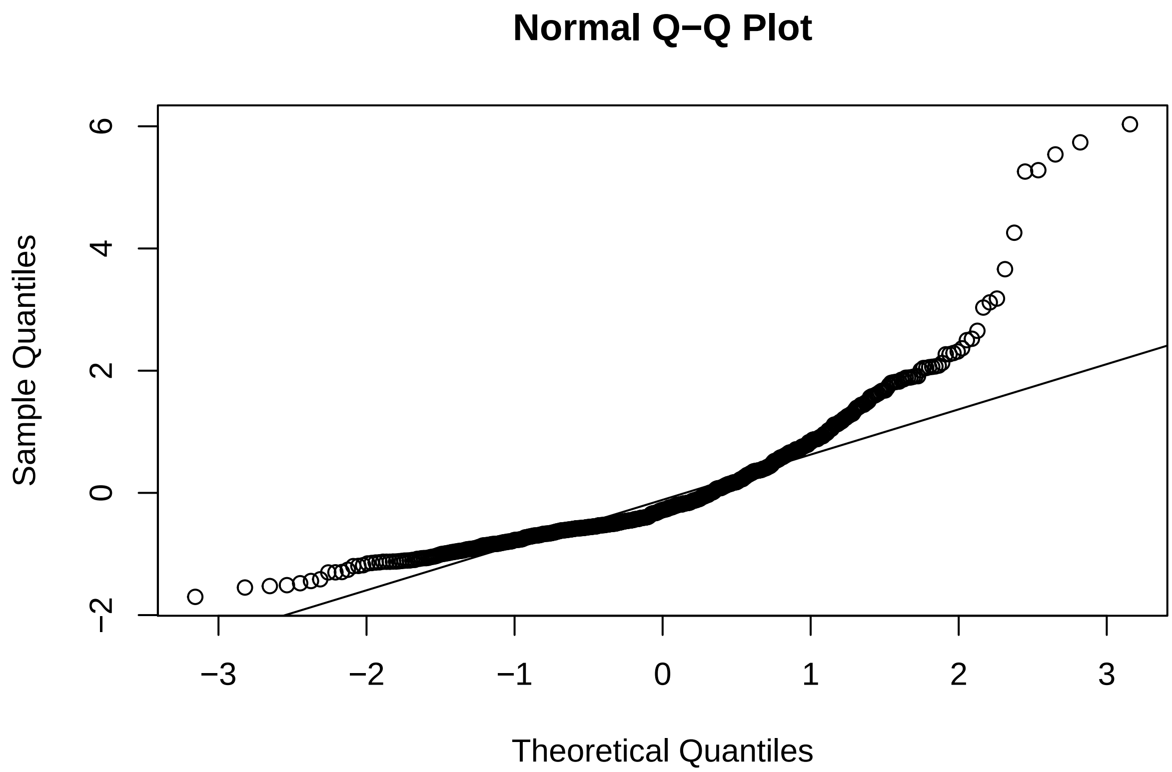
```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    39.0673     0.4080  95.754   <2e-16 ***
## fl.df4$RelEv   -1.7062     0.1658 -10.293   <2e-16 ***
## cosvar         -0.3355     0.2690  -1.247    0.213
## sinvar         -0.1298     0.2745  -0.473    0.637
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.726 on 623 degrees of freedom
## Multiple R-squared:  0.1526, Adjusted R-squared:  0.1485
## F-statistic: 37.39 on 3 and 623 DF,  p-value: < 2.2e-16
```





Diagnostics of extended cosine model, no outliers:

```
##
##  Shapiro-Wilk normality test
##
## data:  delay4resid
## W = 0.83522, p-value < 2.2e-16
```




```
##
##  Bartlett test of homogeneity of variances
##
## data:  delay4resid and fl.df4$Rel.hrDay
## Bartlett's K-squared = 382.02, df = 32, p-value < 2.2e-16
```

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  delay4resid and fl.df4$Rel.hrDay
## Fligner-Killeen:med chi-squared = 160.68, df = 32, p-value <
## 2.2e-16
```

Hm. Looks better, closer to homoscedasticity of residuals and no points with excessive leverage, but normality of residuals still seems pretty far off.

If we don't trust the linear modeling approach, we can try the Jammalamadaka-Sarma (circular) correlation coefficient to assess the reltionship between release time and arrival time (derivatives of transit time)

This is a circular version of the PEarson's product moment correlation, and programed into the 'circular' package

```r
relhrcirc = as.circular(fl.df2$Rel.hrDay, units="hours", template="clock24", rotation="clock")
```

```
## Warning in as.circular(fl.df2$Rel.hrDay, units = "hours", template = "clock24", : an object is coerc
##   type: 'angles'
##   modulo: 'asis'
##   zero: 0
## evalexprenvirenclos
```

```r
F.hrcirc = as.circular(fl.df2$F.hrDay, units="hours", template="clock24", rotation="clock")
```

```
## Warning in as.circular(fl.df2$F.hrDay, units = "hours", template = "clock24", : an object is coerced
##   type: 'angles'
##   modulo: 'asis'
##   zero: 0
## evalexprenvirenclos
```

```r
  cor.circular(F.hrcirc, relhrcirc, test=T)
```

```
## $cor
## [1] 0.1858327
##
## $statistic
## [1] 4.919331
##
## $p.value
## [1] 8.684036e-07
```

And then use the non-parametric Kruskal-Wallis rank sum test to assess the relationship between transit time (lienar continuous var) and release event (linear factor).

Often can use the Nemenyi test for multiple comparisons after the kruskal test, but Zar notes that this is only appropriate with equal sample sizes across groups, and we have ~200 in Rels 1 & 2, 131 in Rel 3, and ~40 in Rels 4 & 5. Use the Dunn test instead

```r
kruskal.test(fl.df2$Delay.hr~factor(fl.df2$RelEv))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  fl.df2$Delay.hr by factor(fl.df2$RelEv)
## Kruskal-Wallis chi-squared = 187.54, df = 4, p-value < 2.2e-16
```

```r
  # posthoc test:
  library(FSA)
```

```
## Warning: package 'FSA' was built under R version 3.3.2
```

```
##
##
```

```
## ############################################
## ##      FSA package, version 0.8.10       ##
## ##    Derek H. Ogle, Northland College    ##
## ##                                        ##
## ## Run ?FSA for documentation.            ##
## ## Run citation('FSA') for citation ...   ##
## ##   please cite if used in publication.  ##
## ##                                        ##
## ## See derekogle.com/fishR/ for more      ##
## ##   thorough analytical vignettes.       ##
## ############################################
```

```
##
## Attaching package: 'FSA'
```

```
## The following object is masked from 'package:psych':
##
##     headtail
```

```r
dunnTest(fl.df2$Delay.hr ~ factor(fl.df2$RelEv), method="bh")
```

```
## Dunn (1964) Kruskal-Wallis multiple comparison
```

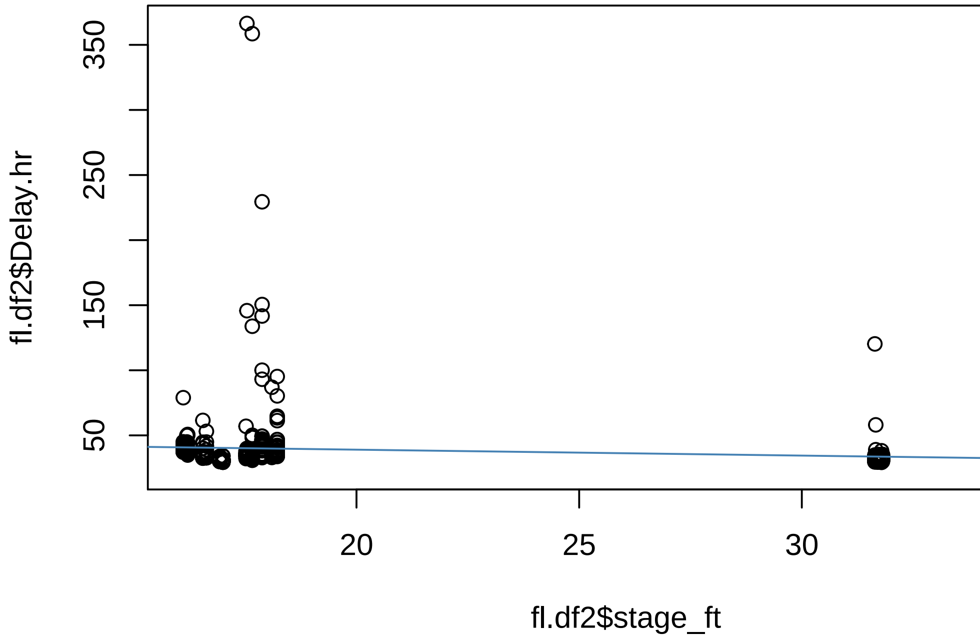
```
##   p-values adjusted with the Benjamini-Hochberg method.
```

```
##    Comparison          Z      P.unadj        P.adj
## 1       1 - 2  3.6041217 3.132103e-04 4.474432e-04
## 2       1 - 3 11.9429117 7.070400e-33 7.070400e-32
## 3       2 - 3  8.7998122 1.370453e-18 6.852263e-18
## 4       1 - 4  7.8800783 3.271760e-15 1.090587e-14
## 5       2 - 4  5.8823517 4.044778e-09 8.089556e-09
## 6       3 - 4  0.2656691 7.904941e-01 7.904941e-01
## 7       1 - 5  5.9571274 2.567101e-09 6.417753e-09
## 8       2 - 5  3.9385898 8.196191e-05 1.366032e-04
## 9       3 - 5 -1.6448327 1.000043e-01 1.250054e-01
## 10      4 - 5 -1.5357427 1.246015e-01 1.384461e-01
```

We see that releases 3, 4, and 5 are not different in thier delay times, but rel 1 and 2 are each unique.

The release events appear correlated with river stage - does stage a better predictor of the differences between groups?

- this code tries to use linear regression, but both the delay times and the stage measurements are dreadfully non-normal. Need to find another test, if we'd like to use a statistical test. Again, leaving this incomplete



until I know if it will be of interest.

```
##
## Call:
## lm(formula = fl.df2$Delay.hr ~ fl.df2$stage_ft)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.00  -5.69  -2.82  -0.29 326.28
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       48.0864     2.7745  17.332  < 2e-16 ***
## fl.df2$stage_ft  -0.4541     0.1172  -3.874 0.000118 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.64 on 639 degrees of freedom
## Multiple R-squared:  0.02295,    Adjusted R-squared:  0.02142
## F-statistic: 15.01 on 1 and 639 DF,  p-value: 0.0001179
```

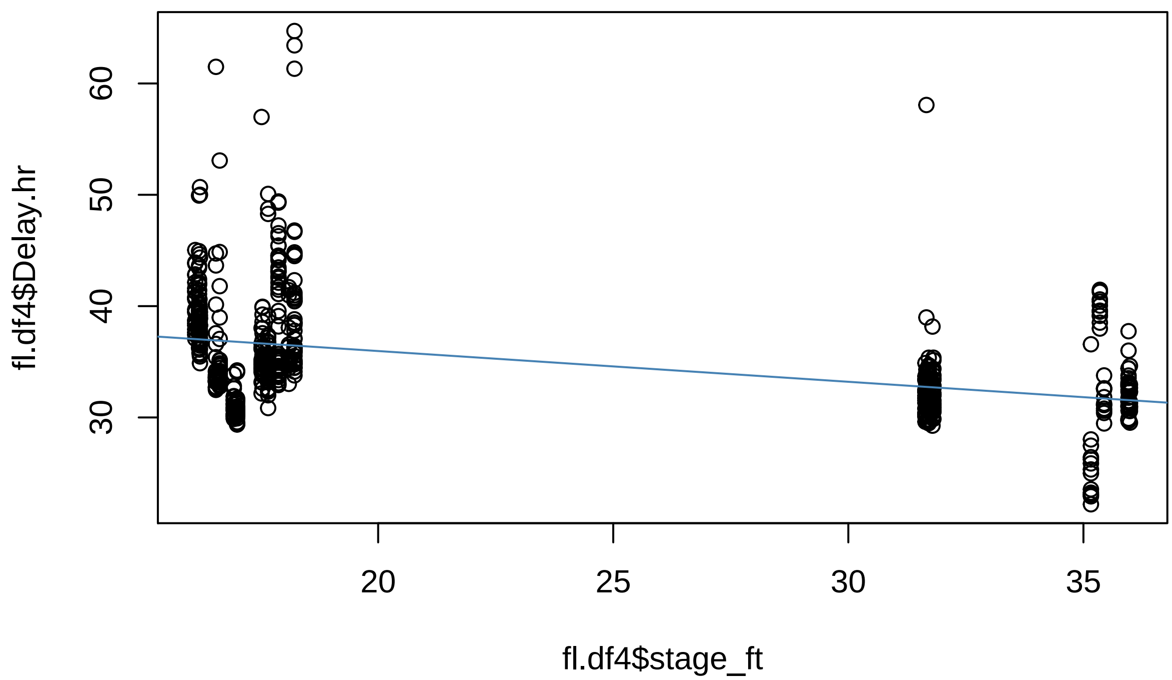
At higher stages fish actually took longer. Is this driven by the heavy outliers?

- Use the reduced dataset from above without the longest 12 travel times (>72)

```
fl.df4 = fl.df2[fl.df2$Delay.hr<72,]

plot(fl.df4$Delay.hr ~ fl.df4$stage_ft)
```

```r
stgmod = lm(fl.df4$Delay.hr ~ fl.df4$stage_ft)
abline(stgmod, col="steelblue")
```

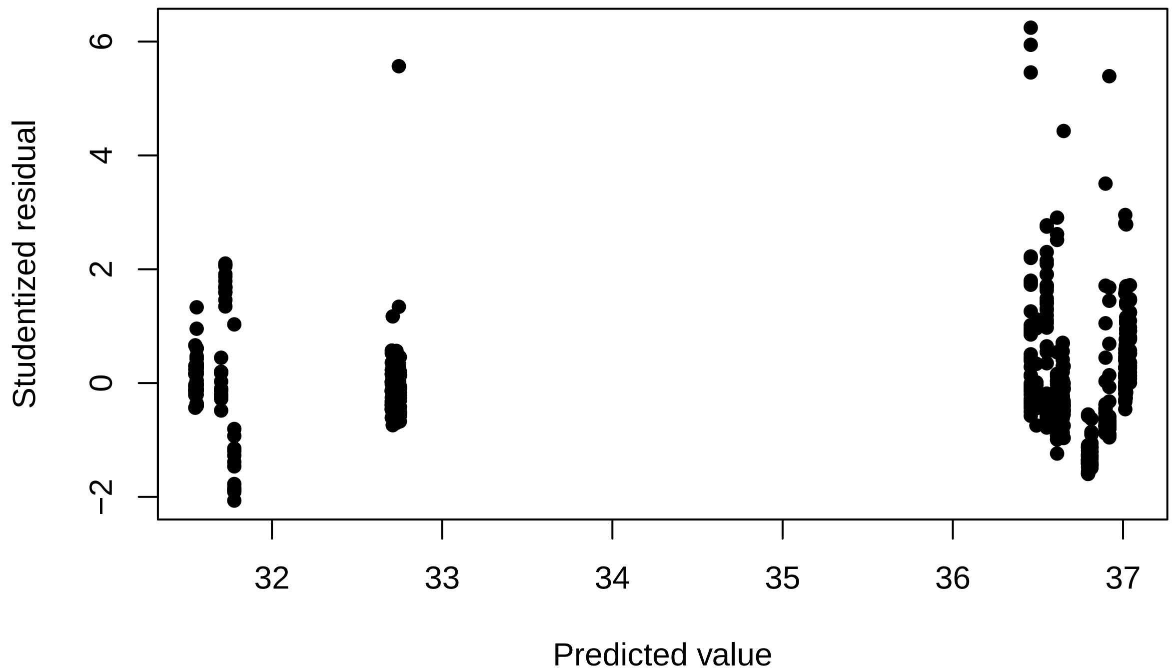



```r
summary(stgmod)
```

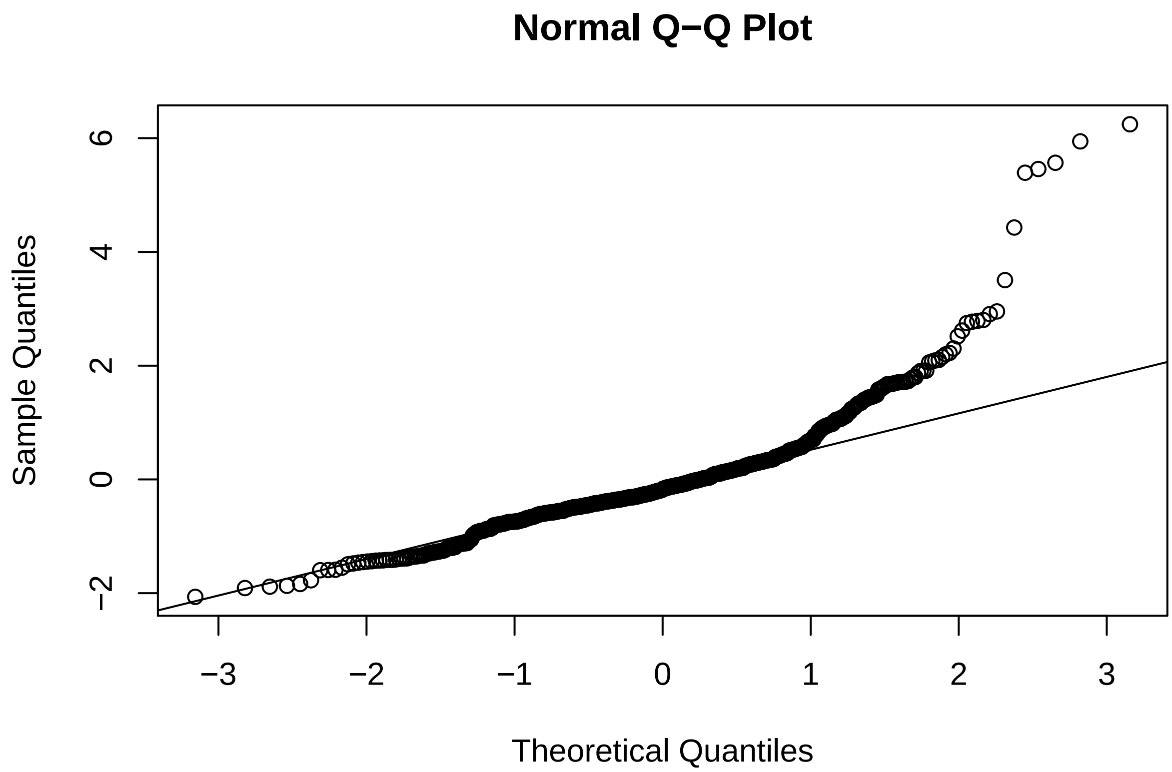
```
##
## Call:
## lm(formula = fl.df4$Delay.hr ~ fl.df4$stage_ft)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5767 -2.5701 -0.7989  1.4577 28.2520
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     41.49055    0.57678   71.94   <2e-16 ***
## fl.df4$stage_ft -0.27622    0.02427  -11.38   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.665 on 625 degrees of freedom
## Multiple R-squared:  0.1717, Adjusted R-squared:  0.1703
## F-statistic: 129.5 on 1 and 625 DF,  p-value: < 2.2e-16
```

```r
#plot(stgmod)
```

Diagnostics




```
##
##  Shapiro-Wilk normality test
##
## data:  stgresid
## W = 0.85669, p-value < 2.2e-16
```



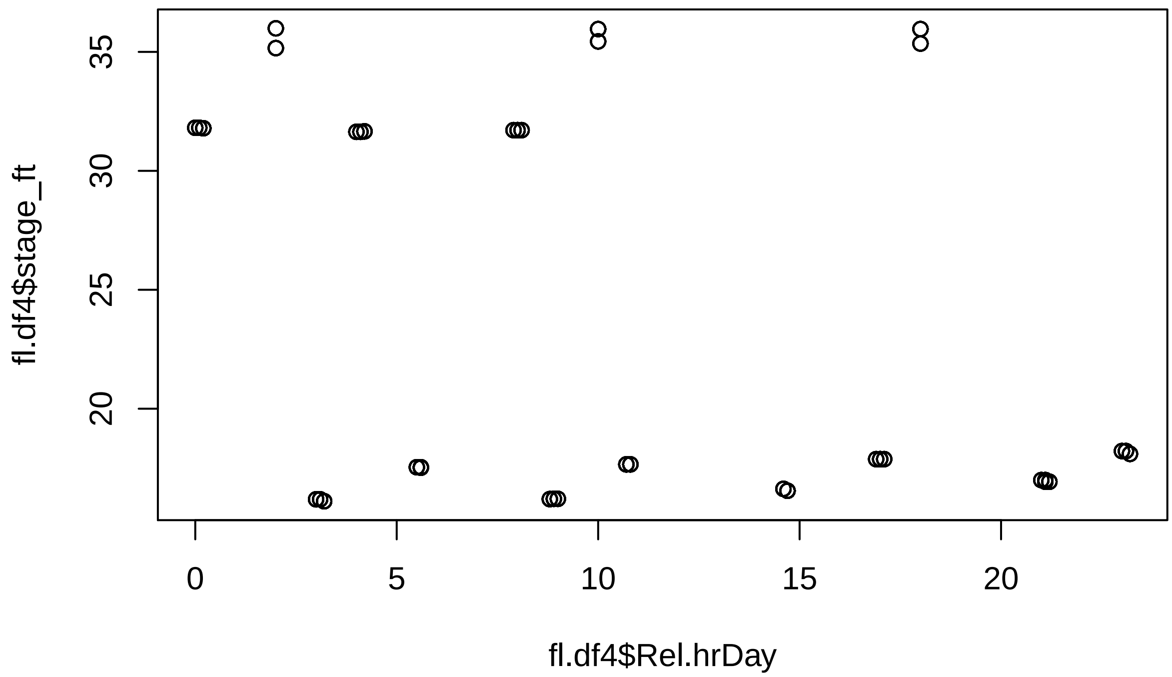

```
##
##  Bartlett test of homogeneity of variances
##
## data:  stgresid and fl.df4$stage_ft
## Bartlett's K-squared = 423.84, df = 23, p-value < 2.2e-16
```

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  stgresid and fl.df4$stage_ft
## Fligner-Killeen:med chi-squared = 158.24, df = 23, p-value <
## 2.2e-16
```

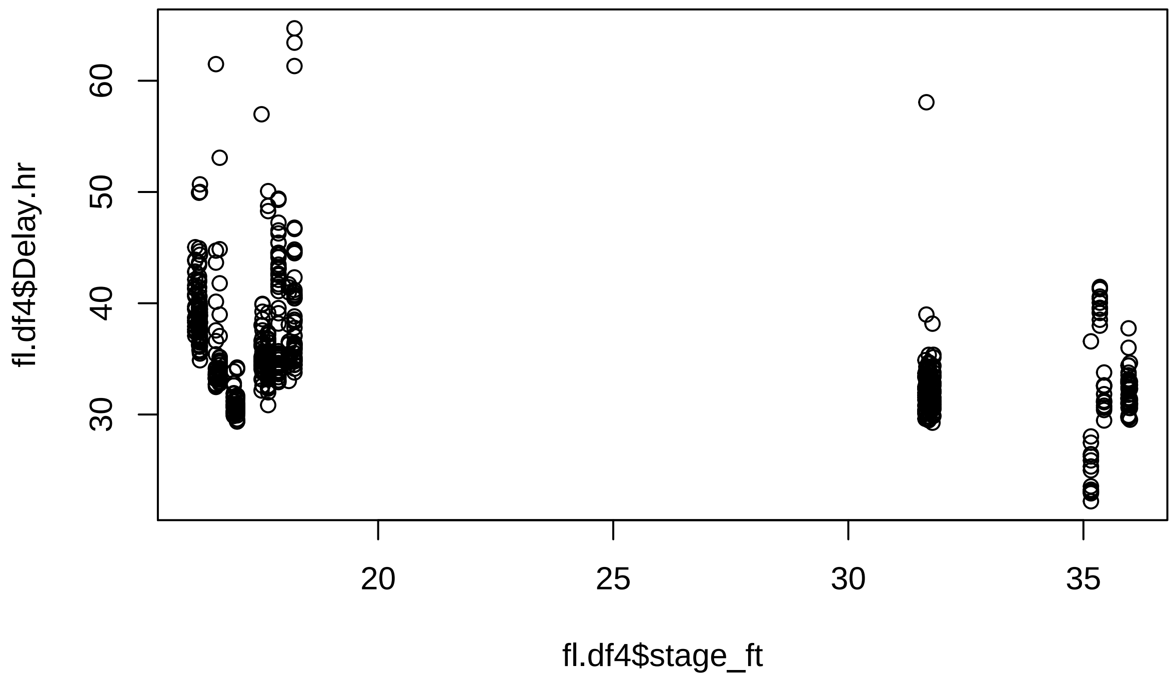
The fit is very significant with the releases 3, 4, & 5 ($p<0.0001$) whereas without these releases it had been ($p=0.83$). But the diagnistic plots still don't look good, primarily the normality of residuals.

One final set of plots to simply look at the relationships in data:

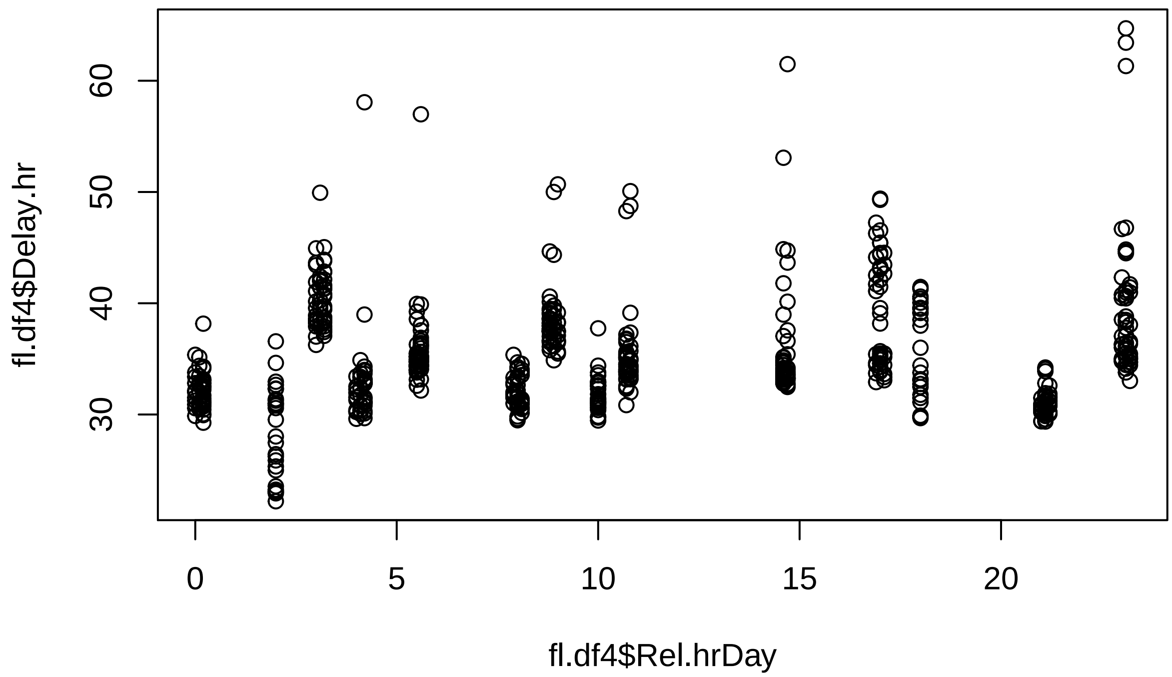
```
plot(fl.df4$stage_ft ~ fl.df4$Rel.hrDay)
```




```
plot(fl.df4$Delay.hr ~ fl.df4$stage_ft)
```




```
plot(fl.df4$Delay.hr ~ fl.df4$Rel.hrDay)
```



Based on the plots and the models (both those that that violate assumptions and the less powerful non-parametric models), the big picture that emerges is that the fish take ~30-50 hours to transit the 55.1 km between the release and the array, and this doesn't change based on the time of day they enter the river. But there is an important effect of release event on transit time, which is most likely driven by the differences in river stage across release events. Therefore, I will continue with the analysis, ignoring the time of release but including the release event as a covariate (unless it is colinear with stage/dischage).