Recasting Galant Schema Statistics

Prepared by Nat Condit-Schultz

October 9, 2018

Loading Data

I first load the data.table package, and then import the data as a data.table.

```
library(data.table)

setwd("~/Bridge/Work/Statistics/AaronCarterEnyi/SchemaCounterPointPaper")
allOnsets <- fread('GR_ACE_StatsData.csv')
```

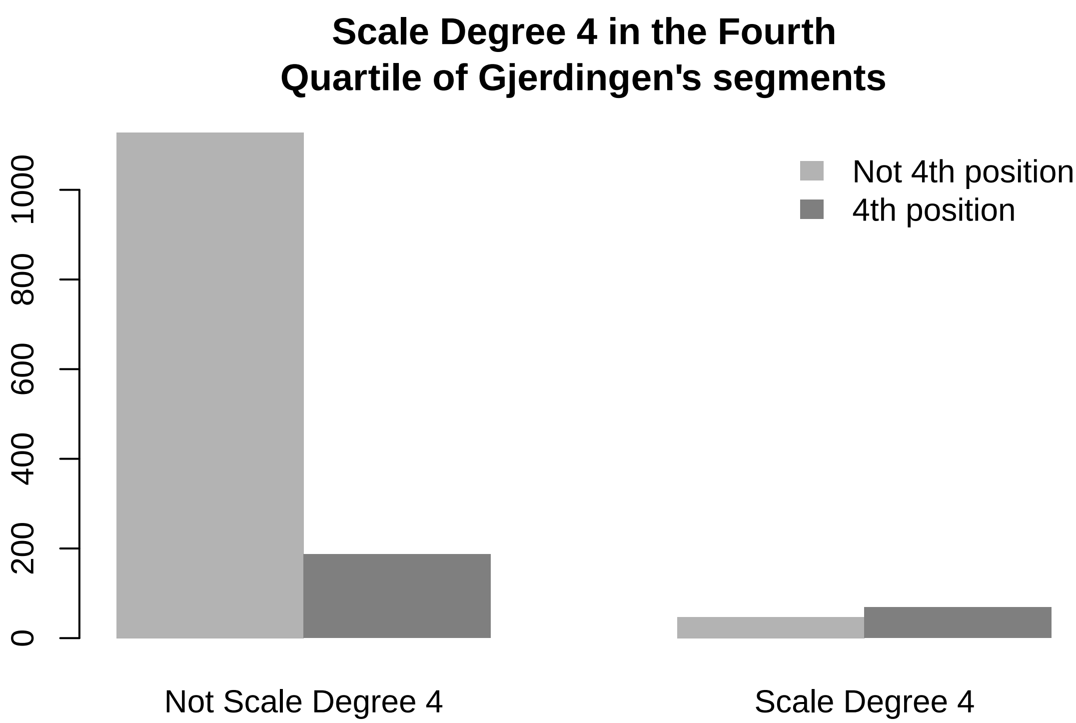
To start, I will take the *first* onset appearence for each note. I create a newtable which is just the firsts.

```
firstOnsets <- allOnsets[ , lapply(.SD,
                                   function(col) {
                                     nobrackets <- gsub('[^1-9,]', '', col)
                                     split <- strsplit(nobrackets, split = ',')
                                     as.numeric(sapply(split, function(str) str[1]))
                                     })]
```

Analysis one

We are trying to show that scale degree four is more likely to occur later in the window. For now, I will just do Gjergingen's segments. The simplest way to visualize this would be a 2x2 table like this:

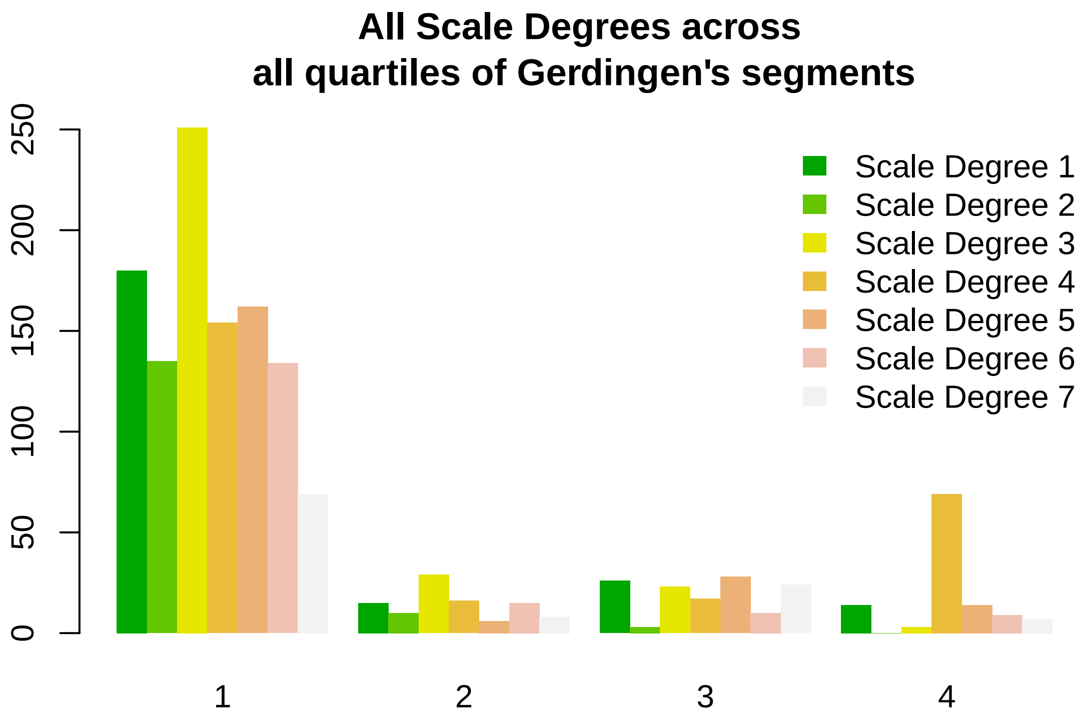
```
twoBytwo <- function(var, main = '') {

  firstOnsets[ , barplot(table(ScaleDegree == 4, eval(var) == 4), beside = TRUE, border = NA,
                         col = c('grey70', 'grey50'), main = main,
                         names.arg = c('Not Scale Degree 4', 'Scale Degree 4'))] -> x

  legend('topright', legend = c('Not 4th position', '4th position'), border = NA,
         fill = c('grey70', 'grey50'),
         bty= 'n')
}

twoBytwo(quote(MetricSegment), "Scale Degree 4 in the Fourth\nQuartile of Gjerdingen's segments")
```



for most scale degrees, the fourth rhythmic position is *much* rarer than the other positions, but for scale degree four, the fourth position is actually more common.

If we tabulate all the data, there doesn't seem to be any other patterns:

```
allBars <- function(var, main = '') {

  firstOnsets[ , barplot(table(ScaleDegree, eval(var)), beside = TRUE, border = NA,
                         main = main,
                         col = terrain.colors(7))] -> x

  legend('topright', legend = paste('Scale Degree', 1:7), border = NA,
         fill = terrain.colors(7),
         bty= 'n')
}

allBars(quote(MetricSegment), "All Scale Degrees across \nall quartiles of Gerdingen's segments")
```



Indeed it seems that scale degree four in segment four is the real outlier!

The model is essentially a T-test, except I'll use a logistic model, since our data are binary hits (either it is scale degree four, or it isn't.)

```
stats <- function(var) {

firstOnsets[!is.na(ScaleDegree) & !is.na(eval(var)) ,
            glm(eval(var) == 4 ~ 1, family = binomial())] -> null
firstOnsets[!is.na(ScaleDegree) & !is.na(eval(var)) ,
            glm(eval(var) == 4 ~ ScaleDegree == 4, family = binomial())] -> fit

print(anova(null, fit, test = 'Chisq'))

print(summary(fit))

return(summary(fit))

}

mod <- stats(quote(MetricSegment))
```

```
## Analysis of Deviance Table
##
## Model 1: eval(var) == 4 ~ 1
## Model 2: eval(var) == 4 ~ ScaleDegree == 4
##   Resid. Df Resid. Dev Df Deviance              Pr(>Chi)
## 1      1430     805.24
## 2      1429     693.06  1   112.18 < 0.00000000000000022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Call:
## glm(formula = eval(var) == 4 ~ ScaleDegree == 4, family = binomial())
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7925  -0.2857  -0.2857  -0.2857   2.5373
##
## Coefficients:
##                      Estimate Std. Error z value            Pr(>|z|)
## (Intercept)           -3.1781     0.1489  -21.35 <0.0000000000000002 ***
## ScaleDegree == 4TRUE   2.1811     0.2049   10.64 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 805.24  on 1430  degrees of freedom
## Residual deviance: 693.06  on 1429  degrees of freedom
## AIC: 697.06
##
## Number of Fisher Scoring iterations: 6
```

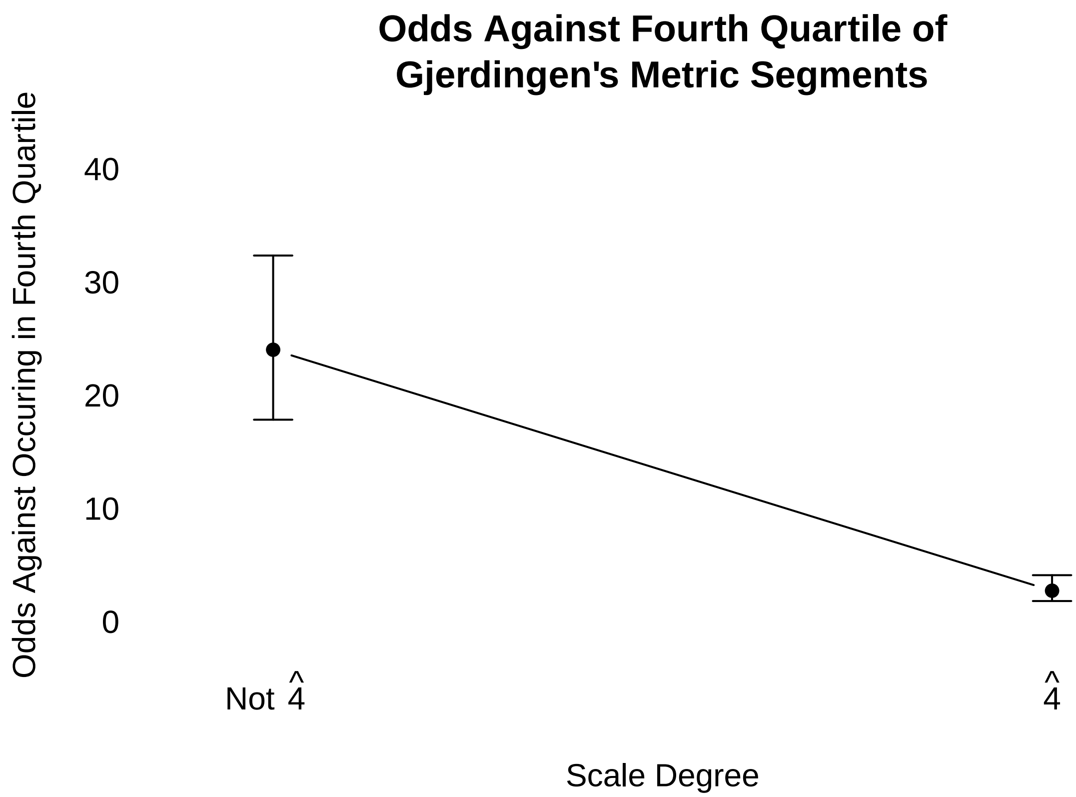
```
coefs <- coef(mod)
```

The model is significant at $\alpha = .05$. You could report the regression coefficient ($\beta = 10.6, Df = 1, p < .05$) or the log-likelihood ratio test ($\chi^2 = 224.36, Df = 1, p < .05$).

For most scale degrees the odds of occuring in the fourth metric position are about 1 to 24. However, for the fourth scale degree, the odds of occuring in the fouth metric position are 8.85 times greater, at about 1 to 2.7. The 95% confidence limit on that estimate of 8.85, ranges from 5.87 to 13.34.

```
odds <- function(x) 1 / (exp(x))

oddsPlots <- function(var, main = '') {

  ytop <- odds(min(coefs[1,1] - coefs[1,2], coefs[2,1] - coefs[2,2]))*1.5

  plot(1:2, odds(coefs[1,1] + c(0, coefs[2,1])), type = 'b', pch = 16,
       ylim = c(0, ytop), axes = FALSE, xlim = c(.9, 2.1),
       main = main,
       xlab = 'Scale Degree', ylab = 'Odds Against Occuring in Fourth Quartile')
  axis(1, .97, 'Not', tick = FALSE)
  axis(1, 1.03, quote(hat(4)), tick = FALSE)
  axis(1, 2, quote(hat(4)), tick = FALSE)

  axis(2, pretty(0:ytop), las = 1, tick = FALSE)

  arrows(x0 = 1, x1 = 1,
         y0 =  odds(coefs[1,1] + 2*coefs[1,2]),
         y1 = odds(coefs[1,1] - 2*coefs[1,2]),
         code = 3, angle = 90, length =.1)

  arrows(x0 = 2, x1 = 2,
         y0 =  odds(sum(coefs[,1]) + 2*coefs[2,2]),
         y1 = odds(sum(coefs[,1]) - 2*coefs[2,2]),
```

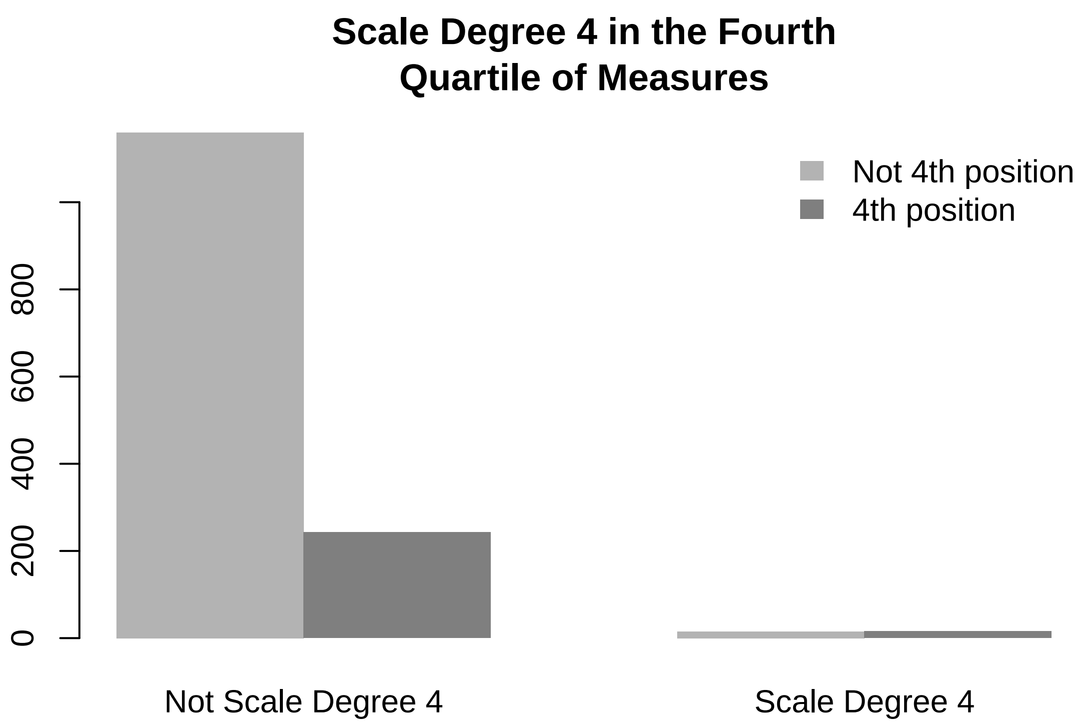
```r
        code = 3, angle = 90, length =.1)
}

oddsPlots(quote(MetricSegment), main = "Odds Against Fourth Quartile of\nGjerdingen's Metric Segments")
```



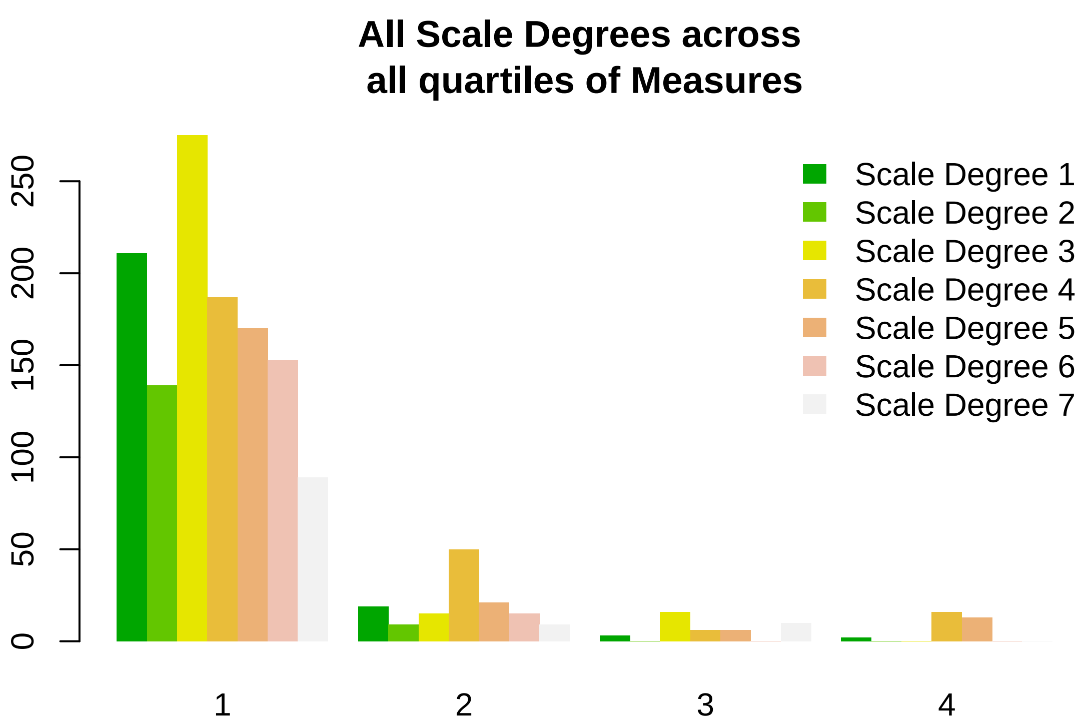
Analysis 2 and 3

I reproduce the analyses, but using the other two measures. First, measures:

```r
twoBytwo(quote(Measure), "Scale Degree 4 in the Fourth\nQuartile of Measures")
```




```
allBars(quote(Measure), "All Scale Degrees across \nall quartiles of Measures")
```

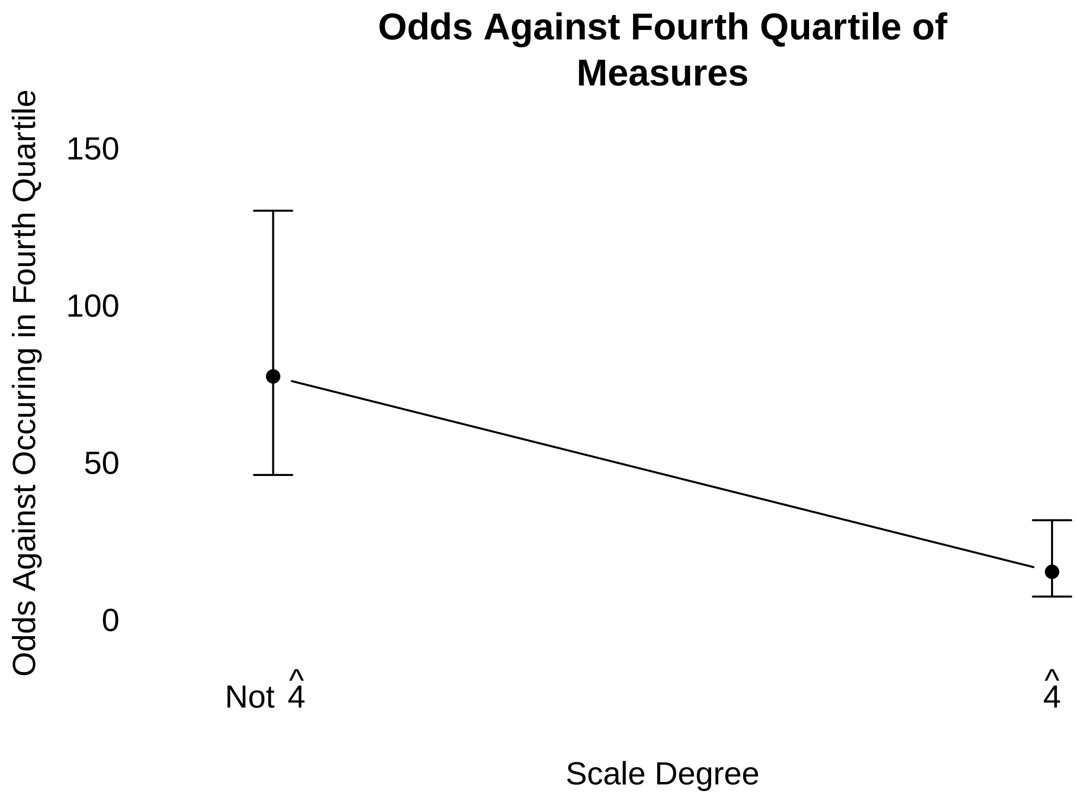



```
mod <- stats(quote(Measure))
```

```
## Analysis of Deviance Table
##
## Model 1: eval(var) == 4 ~ 1
## Model 2: eval(var) == 4 ~ ScaleDegree == 4
##   Resid. Df Resid. Dev Df Deviance   Pr(>Chi)
```

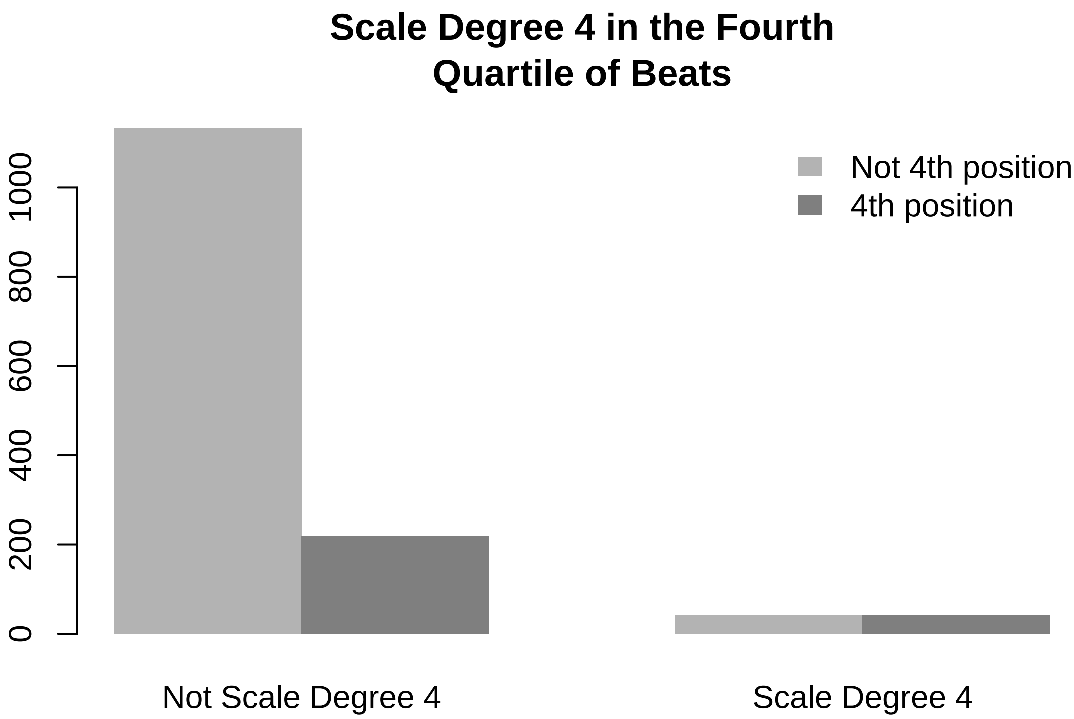
```
## 1      1433     299.05
## 2      1432     280.72  1   18.325 0.00001863 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Call:
## glm(formula = eval(var) == 4 ~ ScaleDegree == 4, family = binomial())
##
## Deviance Residuals:
##     Min        1Q    Median       3Q      Max
## -0.3571  -0.1603  -0.1603  -0.1603   2.9533
##
## Coefficients:
##                      Estimate Std. Error z value             Pr(>|z|)
## (Intercept)           -4.3481     0.2599 -16.732 < 0.0000000000000002 ***
## ScaleDegree == 4TRUE   1.6277     0.3663   4.444           0.00000883 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 299.05  on 1433  degrees of freedom
## Residual deviance: 280.72  on 1432  degrees of freedom
## AIC: 284.72
##
## Number of Fisher Scoring iterations: 7
```

```
coefs <- coef(mod)
oddsPlots(quote(Measure), main = "Odds Against Fourth Quartile of\nMeasures")
```

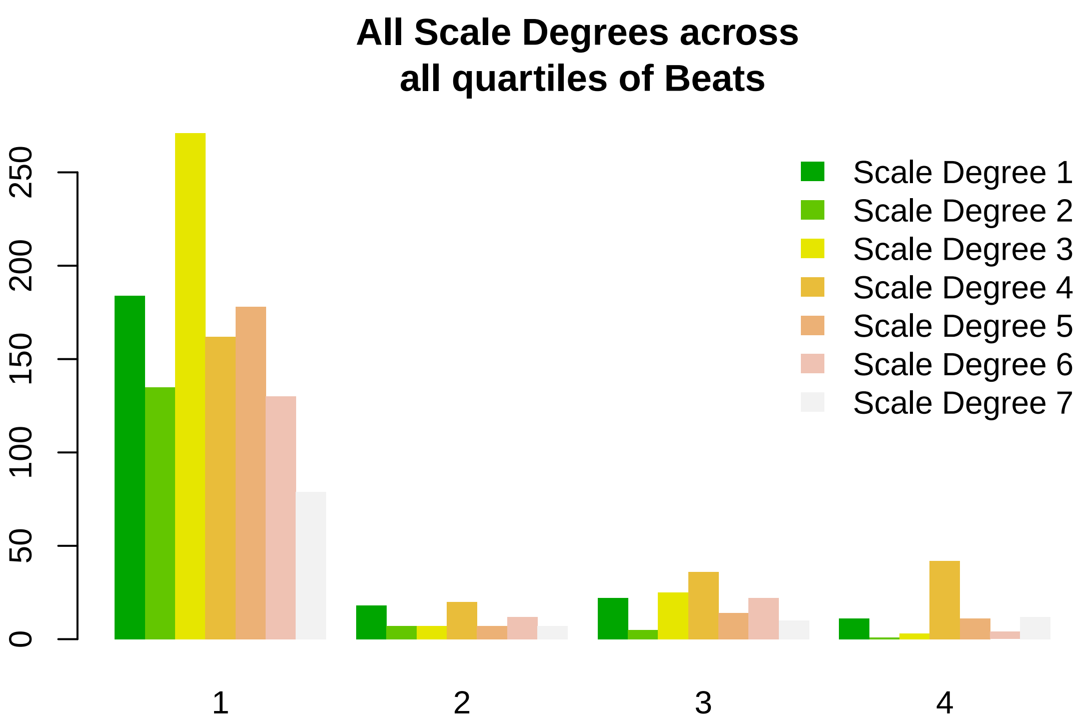


Then beats:

```
twoBytwo(quote(Beat), "Scale Degree 4 in the Fourth\nQuartile of Beats")
```




```
allBars(quote(Beat), "All Scale Degrees across \nall quartiles of Beats")
```

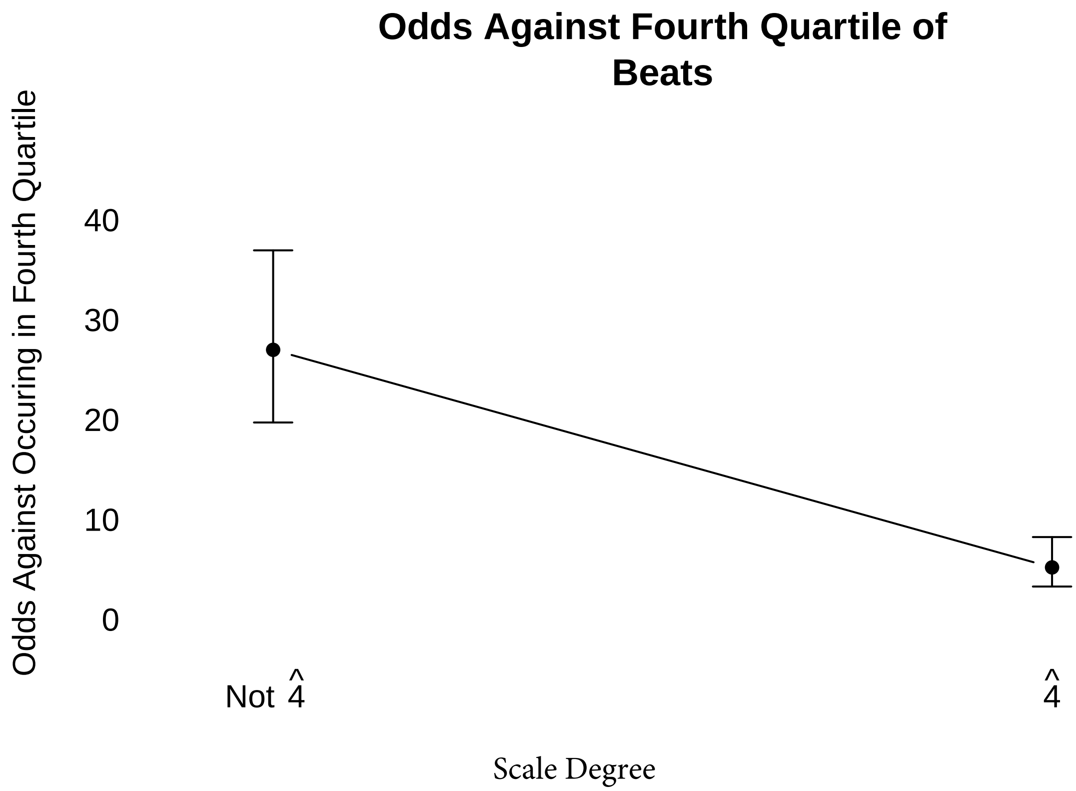



```
mod <- stats(quote(Beat))
```

```
## Analysis of Deviance Table
##
```

```
## Model 1: eval(var) == 4 ~ 1
## Model 2: eval(var) == 4 ~ ScaleDegree == 4
##   Resid. Df Resid. Dev Df Deviance          Pr(>Chi)
## 1      1434     639.79
## 2      1433     592.26  1   47.521 0.000000000005442 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Call:
## glm(formula = eval(var) == 4 ~ ScaleDegree == 4, family = binomial())
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5936  -0.2698  -0.2698  -0.2698   2.5812
##
## Coefficients:
##                      Estimate Std. Error z value             Pr(>|z|)
## (Intercept)           -3.2950     0.1571 -20.969 < 0.0000000000000002 ***
## ScaleDegree == 4TRUE   1.6481     0.2304   7.153    0.000000000000849 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 639.79  on 1434  degrees of freedom
## Residual deviance: 592.26  on 1433  degrees of freedom
## AIC: 596.26
##
## Number of Fisher Scoring iterations: 6
```

```r
coefs <- coef(mod)
oddsPlots(quote(Beat), main = "Odds Against Fourth Quartile of\nBeats")
```



In the case of measures, it looks like the second quartile also gets relatively more scale degree 4—clearly Gjerdingen tends to break measures like this into two segments.