

# consulting project

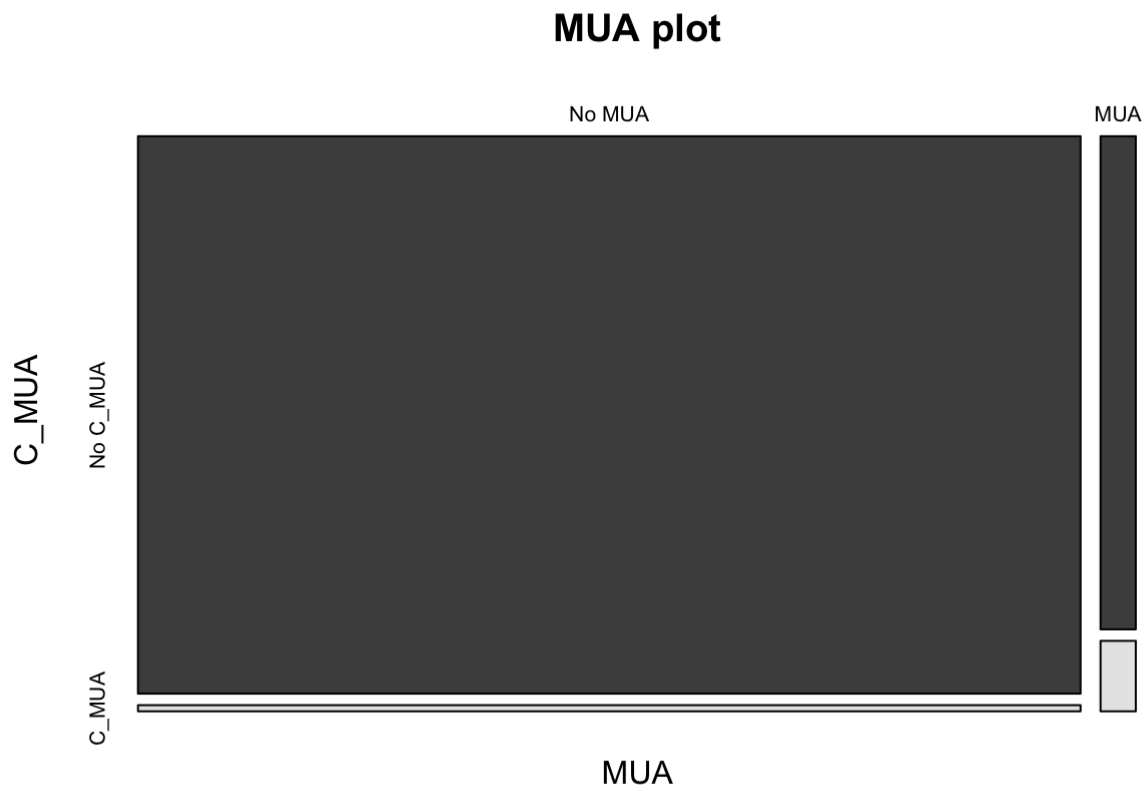
2022-11-10

Q1: MUA is a risk factor of C\_MUA? - Yes

```
#Pearson's Chi-squared test
chisq.test(MUA_data1$MUA, MUA_data1$C_MUA)

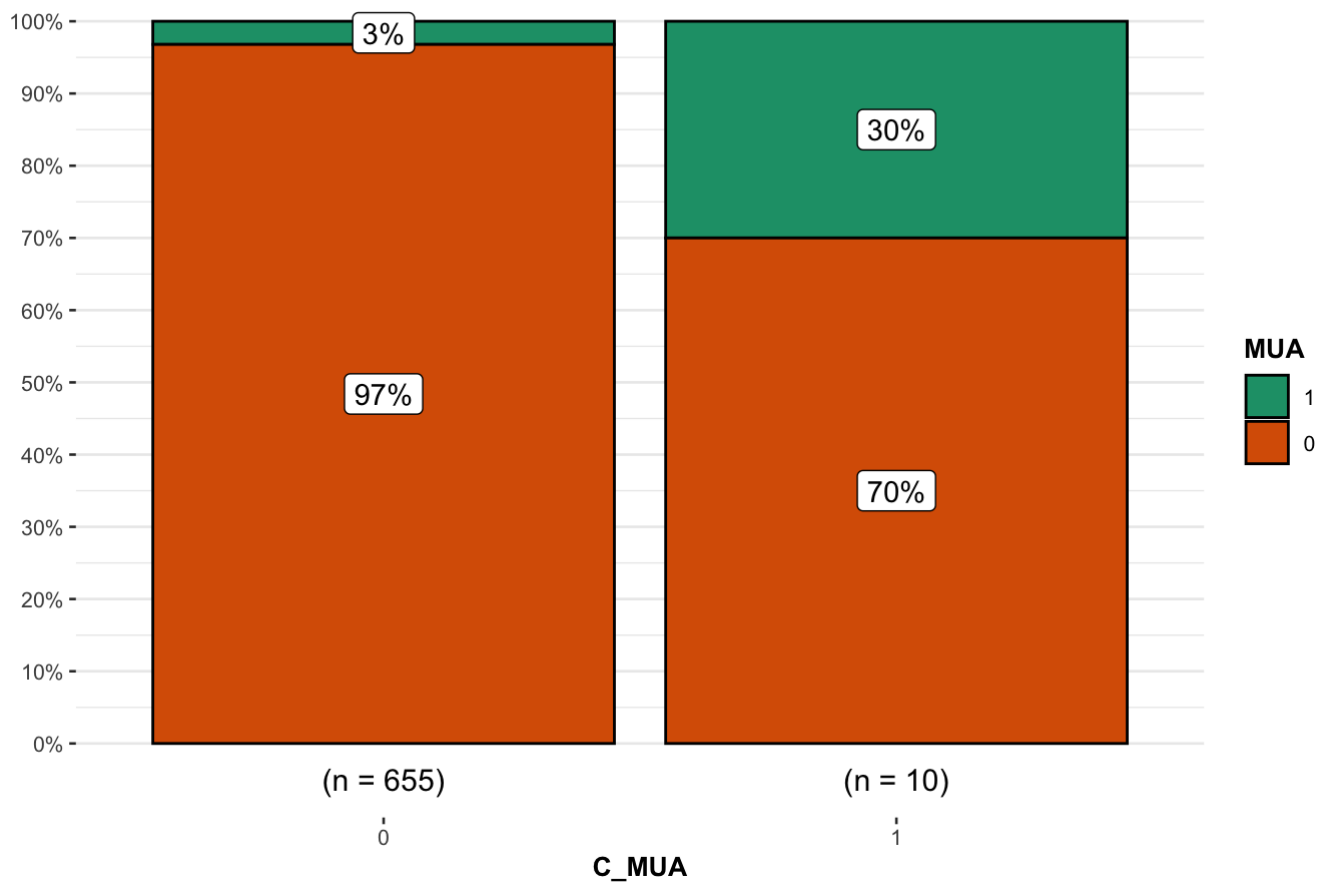
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  MUA_data1$MUA and MUA_data1$C_MUA
## X-squared = 13.354, df = 1, p-value = 0.0002579
```

\* Mosaic plot



\* Fisher's exact test for count data

Fisher's exact test, p-value = 0.004



#### \* Logistic regression model

```
#Model with MUA predicting C_mua  
mua_glm = glm(C_MUA ~ MUA, data= MUA_data1, family = "binomial")  
summary(mua_glm)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5168  -0.1482  -0.1482  -0.1482   3.0057
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.5061     0.3800 -11.857 < 2e-16 ***
## MUAYes        2.5602     0.7248   3.532 0.000412 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  95.248  on 663  degrees of freedom
## AIC: 99.248
##
## Number of Fisher Scoring iterations: 7
```

- $\text{logit}(\text{odds}) = -4.5061 + 2.5602 \cdot \text{MUA\_bi} (0/1)$
- if the patient got MUA, the odds of getting C\_MUA is  $\exp(-4.5061 + 2.5602) = 0.1428586$
- if the patient didn't get MUA, the odds of getting C\_MUA is  $\exp(-4.5061) = 0.01104144$

## \* Odds Ratio & Correlation

```
#Odds ratio
exp(mua_glm$coefficients[-1])
```

```
##      MUAYes
## 12.93878
```

Odds of receiving MUA on contralateral knee is almost 13x times larger than not

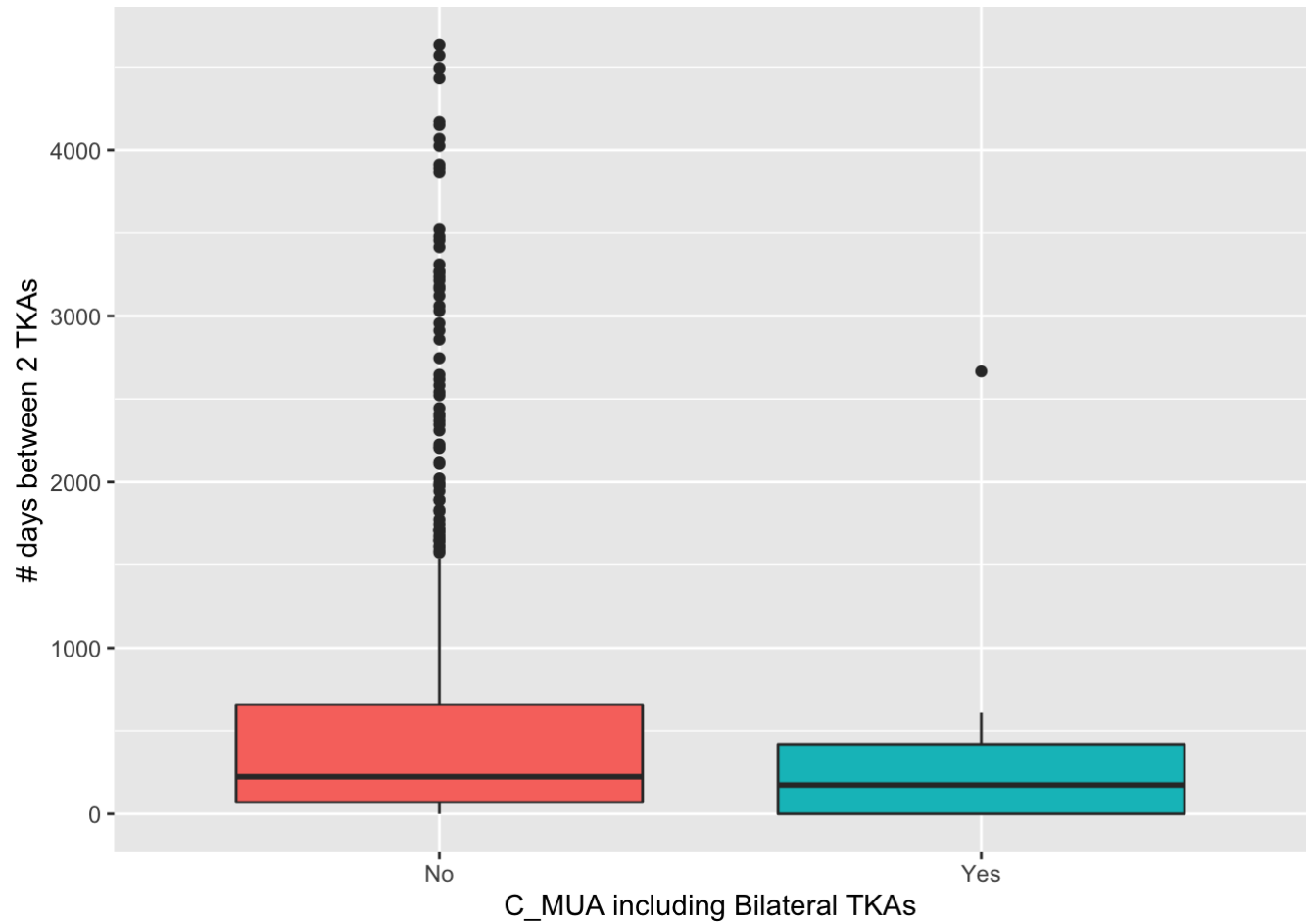
```
cor(MUA_data1$C_MUA_bi, MUA_data1$MUA_bi)
```

```
## [1] 0.1748324
```

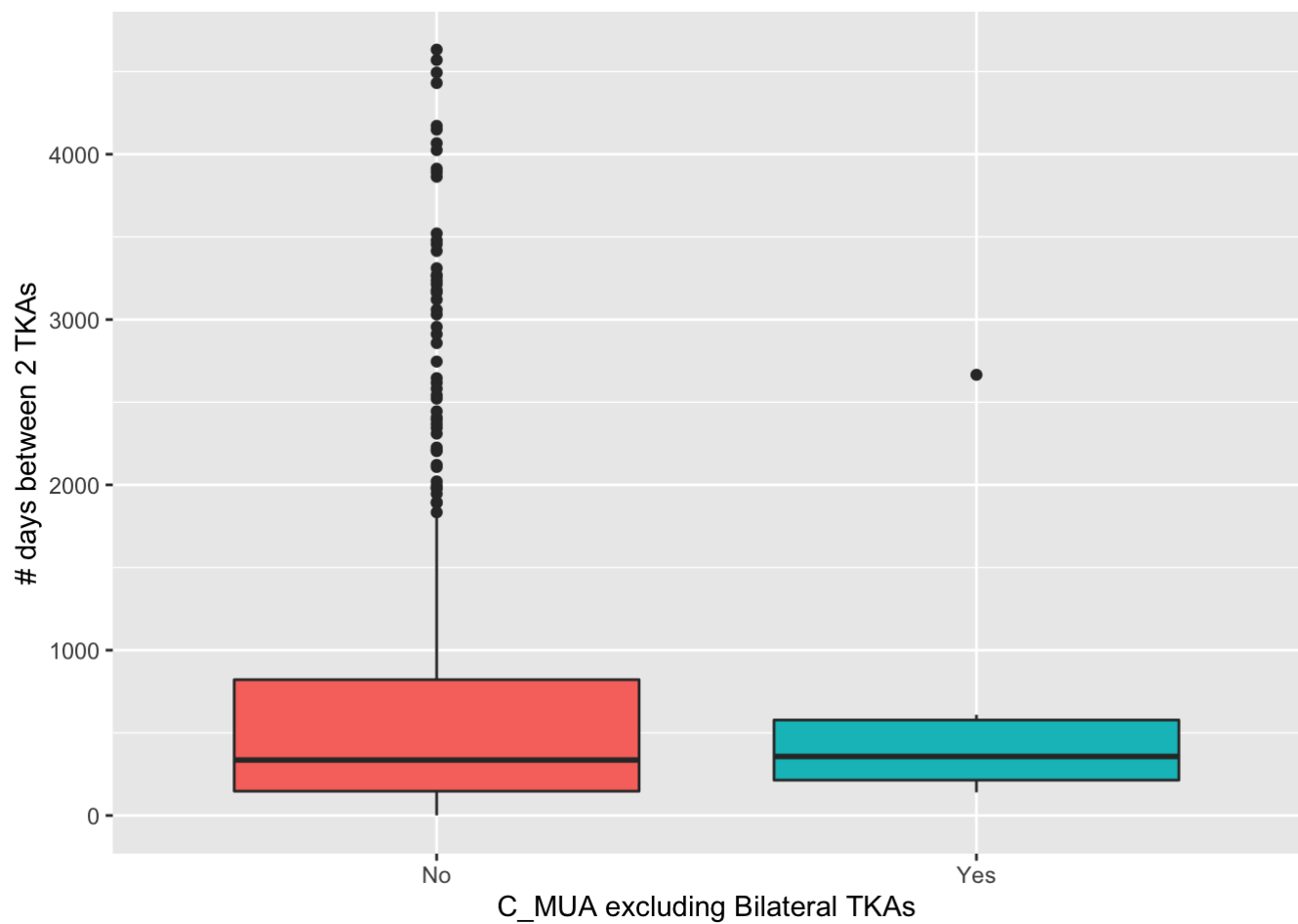
Q2: is the # between 2 TKAs a risk factor? - No

\* boxplot of # days between 2 TKAs

Variable	N	No, N = 655 <sup>1</sup>	Yes, N = 10 <sup>1</sup>
date_diff	665	224 (70, 658)	174 (0, 420)
<sup>1</sup> Median (IQR)			



Variable	N	No, N = 528 <sup>1</sup>	Yes, N = 6 <sup>1</sup>
date_diff	534	336 (147, 822)	357 (214, 578)
<sup>1</sup> Median (IQR)			



```
#Model with MUA predicting C_mua with time  
cor(MUA_data1$C_MUA_bi,MUA_data1$date_diff)
```

```
## [1] -0.01729522
```

```
mua_glm1 = glm(C_MUA ~ date_diff, data= MUA_data1, family = "binomial")  
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ date_diff, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1830  -0.1817  -0.1787  -0.1702   3.0462
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.0812948  0.3773046 -10.817  <2e-16 ***
## date_diff   -0.0002058  0.0004644  -0.443    0.658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 103.57  on 663  degrees of freedom
## AIC: 107.57
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ date_diff, data= MUA_data1[MUA_data1$BTKA=="No",], family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ date_diff, family = "binomial", data = MUA_data1[MUA_data1$BTKA
==
##      "No", ])
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.1635   -0.1506   -0.1491   -0.1485    3.0042
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.508e+00  5.196e-01  -8.676   <2e-16 ***
## date_diff    4.287e-05  4.408e-04   0.097    0.923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 65.796  on 533  degrees of freedom
## Residual deviance: 65.787  on 532  degrees of freedom
## AIC: 69.787
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA +date_diff, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + date_diff, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5990  -0.1599  -0.1549  -0.1420   3.2877
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.3332971  0.4184459 -10.356  < 2e-16 ***
## MUAYes       2.7063127  0.7443088   3.636 0.000277 ***
## date_diff    -0.0004001  0.0004977  -0.804 0.421363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  94.435  on 662  degrees of freedom
## AIC: 100.43
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ (MUA +date_diff)^2, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```



```
##
## Call:
## glm(formula = C_MUA ~ (MUA + date_diff)^2, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8144  -0.1484  -0.1484  -0.1481   3.0104
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.503e+00  4.520e-01  -9.961  < 2e-16 ***
## MUAYes         3.569e+00  9.594e-01   3.720  0.000199 ***
## date_diff      -6.689e-06  4.595e-04  -0.015  0.988385
## MUAYes:date_diff -2.333e-03  2.488e-03  -0.938  0.348382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  92.359  on 661  degrees of freedom
## AIC: 100.36
##
## Number of Fisher Scoring iterations: 7
```

the correlation between days between 2TKAs and C\_MUA is close to 0. In logistic regression model, the days between 2 TKAs is insignificant

let's see patients who got the 2nd TKA within 645 days

```
# let's see patients who got the 2nd TKA within 645 days
summary(as.vector(MUA_data1$date_diff)) #3rd Q : 645
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0      69.0    224.0    552.1   645.0   4633.0
```

```
cor(MUA_data1[MUA_data1$date_diff<645, ]$C_MUA_bi, MUA_data1[MUA_data1$date_diff<645, ]$date_diff)
```

```
## [1] 0.01043714
```

```
mua_glm1 = glm(C_MUA ~ date_diff, data= MUA_data1[MUA_data1$date_diff<645&&MUA_data1$BTKA=="No", ], family = "binomial")
```

```
## Warning in MUA_data1$date_diff < 645 && MUA_data1$BTKA == "No": 'length(x) = 665
## > 1' in coercion to 'logical(1)'

## Warning in MUA_data1$date_diff < 645 && MUA_data1$BTKA == "No": 'length(x) = 665
## > 1' in coercion to 'logical(1)'

## Warning in MUA_data1$date_diff < 645 && MUA_data1$BTKA == "No": 'length(x) = 665
## > 1' in coercion to 'logical(1)'

## Warning in MUA_data1$date_diff < 645 && MUA_data1$BTKA == "No": 'length(x) = 665
## > 1' in coercion to 'logical(1)'
```

```
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ date_diff, family = "binomial", data = MUA_data1[MUA_data1$date
_diff <
##      645 && MUA_data1$BTKA == "No", ])
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1830  -0.1817  -0.1787  -0.1702   3.0462
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.0812948   0.3773046  -10.817  <2e-16 ***
## date_diff   -0.0002058   0.0004644   -0.443    0.658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 103.57  on 663  degrees of freedom
## AIC: 107.57
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA+date_diff+ MUA*date_diff, data= MUA_data1[MUA_data1$date_diff
<645, ], family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + date_diff + MUA * date_diff, family = "binomial",
##      data = MUA_data1[MUA_data1$date_diff < 645, ])
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7534  -0.1604  -0.1574  -0.1550   2.9756
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.4151849   0.5973269   -7.392 1.45e-13 ***
## MUAYes         3.3009333   1.1438643    2.886  0.0039 **
## date_diff      0.0002268   0.0024510    0.093  0.9263
## MUAYes:date_diff -0.0014365   0.0042356   -0.339  0.7345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 90.077  on 497  degrees of freedom
## Residual deviance: 79.460  on 494  degrees of freedom
## AIC: 87.46
##
## Number of Fisher Scoring iterations: 7
```

```
# getting obvious that date_diff is insignificant
```

Even though the correlation between days between 2TKAs and C\_MUA changed from negative to positive, it is still close to 0. When we looked into the days within 3rd quartiles, it is getting obvious that date\_diff is insignificant

- If the patient got Bilateral TKA, is the risk of getting MUA significantly higher? : No.

```
# If the patient got Bilateral TKA, is the risk of getting MUA significantly higher? No.
MUA_data1$BTKA <- ifelse(MUA_data1$date_diff==0, "Yes","No")
mua_glm1 = glm(as.factor(Group2) ~ as.factor(BTKA), data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = as.factor(Group2) ~ as.factor(BTKA), family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -2.556    0.316    0.316    0.316    0.316
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.9724     0.2011  14.783  <2e-16 ***
## as.factor(BTKA)Yes  0.2545     0.4984   0.511    0.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 250.61  on 664  degrees of freedom
## Residual deviance: 250.34  on 663  degrees of freedom
## AIC: 254.34
##
## Number of Fisher Scoring iterations: 6
```

```
mua_glm1 = glm(C_MUA ~ as.factor(BTKA), data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ as.factor(BTKA), family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.2490   -0.1503   -0.1503   -0.1503    2.9962
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -4.4773     0.4106 -10.905  <2e-16 ***
## as.factor(BTKA)Yes  1.0194     0.6530   1.561    0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 101.58  on 663  degrees of freedom
## AIC: 105.58
##
## Number of Fisher Scoring iterations: 7
```

- want to mention about below in the presentation?

```
##
## Call:
## glm(formula = C_MUA ~ yeardiff, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2085  -0.1822  -0.1822  -0.1822   3.1593
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -4.0896     0.3811 -10.731  <2e-16 ***
## yeardiff2yr         0.2719     0.8101   0.336    0.737
## yeardiffmore than 2yrs -0.8940     1.0732  -0.833    0.405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 102.66  on 662  degrees of freedom
## AIC: 108.66
##
## Number of Fisher Scoring iterations: 7
```

```
##
## Classification tree:
## tree(formula = C_MUA ~ MUA + date_diff, data = MUA_data1)
## Number of terminal nodes:  14
## Residual mean deviance:  0.09584 = 62.39 / 651
## Misclassification error rate: 0.01504 = 10 / 665
```

```
##
## Classification tree:
## tree(formula = C_MUA ~ MUA + date_diff, data = MUA_data1[MUA_data1$date_diff <
##      645, ])
## Number of terminal nodes:  13
## Residual mean deviance:  0.1137 = 55.13 / 485
## Misclassification error rate: 0.01807 = 9 / 498
```

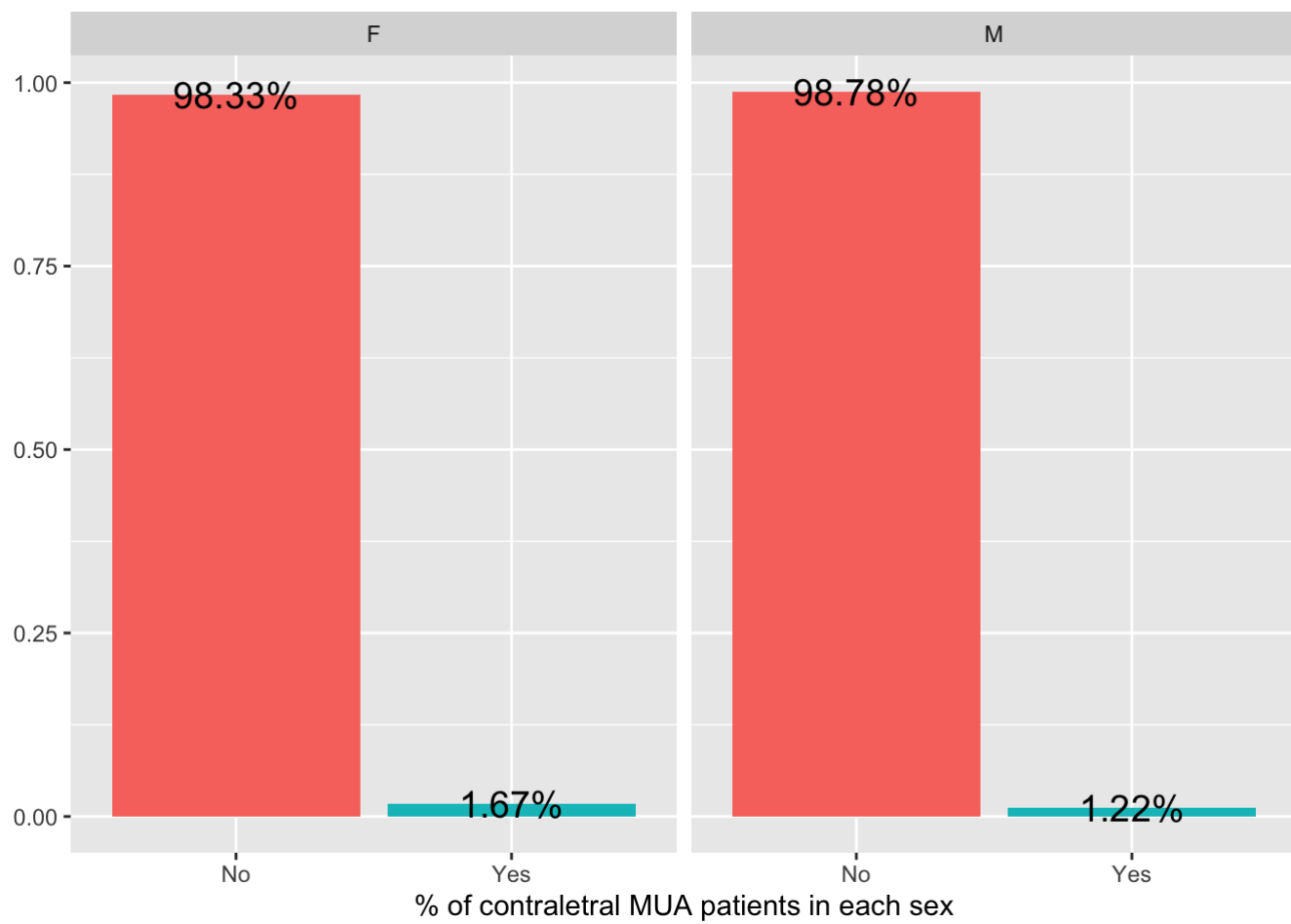
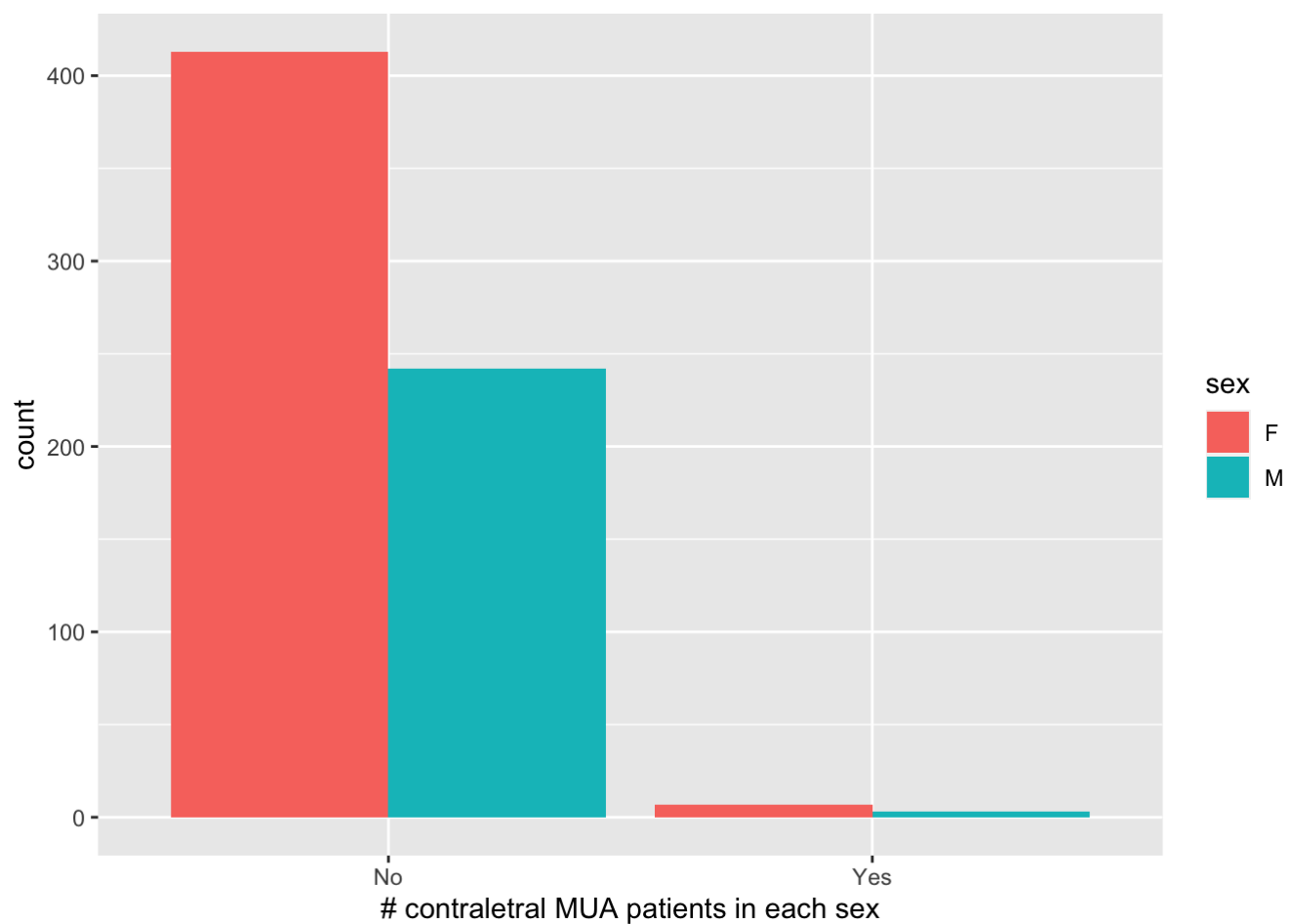
```
##
## Classification tree:
## tree(formula = C_MUA ~ MUA + BTKA, data = MUA_data1)
## Variables actually used in tree construction:
## [1] "MUA"
## Number of terminal nodes:  2
## Residual mean deviance:  0.1437 = 95.25 / 663
## Misclassification error rate: 0.01504 = 10 / 665
```

---

### Q3 Demographics and comorbidities factors?

(1) sex : No

a. EDA



b. chi-squared test

```
names(MUA_data1)
```



```
## [1] "MUA_type"
## [2] "Group2"
## [3] "MUA"
## [4] "C_MUA"
## [5] "ID"
## [6] "sex"
## [7] "age"
## [8] "race"
## [9] "ethnicity"
## [10] "BMI"
## [11] "tobacco"
## [12] "death_date"
## [13] "Insurance"
## [14] "cpt code"
## [15] "surgery_date"
## [16] "Admission date"
## [17] "Discharge date"
## [18] "los"
## [19] "disch_disp"
## [20] "ASA"
## [21] "op_time"
## [22] "mua_count"
## [23] "blood_transfusion"
## [24] "platelet_transfusion"
## [25] "AIDS"
## [26] "Malignancy"
## [27] "Cerebrovascular"
## [28] "COPD"
## [29] "CHF"
## [30] "Dementia"
## [31] "Diabetes_cc"
## [32] "Diabetes_no_cc"
## [33] "Hemiplegia"
## [34] "Metastatic"
## [35] "Mild_Liver"
## [36] "Moderate_Liver"
## [37] "MI"
## [38] "Peptic_Ulcer"
## [39] "PVD"
## [40] "CKD"
## [41] "Rheumatic"
## [42] "hematoma"
## [43] "wound_infection"
## [44] "knee_infection"
## [45] "Readmission within 90 days (1=yes)"
## [46] "Readmission date"
## [47] "Reason for Readmission"
## [48] "ED visit within 90 days (1=yes)"
## [49] "ED visit date"
## [50] "Reason for ED Visit"
## [51] "Initial TKA Knee (right or left)"
## [52] "Pre-Op Total Knee arthroplasty range of motion"
```

```
## [53] "Date of most recent Preop Visit"
## [54] "varus/valgus preoperative (normal=0, varus=1, valgus=2)"
## [55] "Postoperative ROM (list flexion arc) 1"
## [56] "Date of first follow up appt"
## [57] "Postoperative ROM (list flexion arc) 2"
## [58] "Date of second follow up"
## [59] "Date of MUA"
## [60] "IF SAME KNEE HAS A 2ND MUA: Post MUA ROM ARC"
## [61] "DATE of follow up"
## [62] "DATE of 2nd MUA"
## [63] "Date of contralateral TKA"
## [64] "Contralateral side (left or right knee)"
## [65] "Pre-Op contralateral TKA ROM"
## [66] "Date of Most Recent Preop Visit_C_TKA"
## [67] "varus/valgus preoperative_C_TKA"
## [68] "Postoperative ROM (list flexion arc) 1_C_TKA"
## [69] "Date of f/u 1_C_TKA"
## [70] "Postoperative ROM (list flexion arc) 2_C_TKA"
## [71] "Date of f/u 2_C_TKA"
## [72] "Contralateral knee MUA?"
## [73] "Contralateral Knee MUA Date"
## [74] "age_C_TKA"
## [75] "bmi_C_TKA"
## [76] "tobacco_C_TKA"
## [77] "Insurance_C_TKA"
## [78] "los_C_TKA"
## [79] "disch_disp_C_TKA"
## [80] "ASA_C_TKA"
## [81] "op_time_C_TKA"
## [82] "blood_transfusion_C_TKA"
## [83] "platelet_transfusion_C_TKA"
## [84] "AIDS_C_TKA"
## [85] "Malignancy_C_TKA"
## [86] "Cerebrovascular_C_TKA"
## [87] "COPD_C_TKA"
## [88] "CHF_C_TKA"
## [89] "Dementia_C_TKA"
## [90] "Diabetes_cc_C_TKA"
## [91] "Diabetes_no_cc_C_TKA"
## [92] "Hemiplegia_C_TKA"
## [93] "Metastatic_C_TKA"
## [94] "Mild_Liver_C_TKA"
## [95] "Moderate_Liver_C_TKA"
## [96] "MI_C_TKA"
## [97] "Peptic_Ulcer_C_TKA"
## [98] "PVD_C_TKA"
## [99] "CKD_C_TKA"
## [100] "Rheumatic_C_TKA"
## [101] "hematoma_C_TKA"
## [102] "wound_infection_C_TKA"
## [103] "knee_infection_C_TKA"
## [104] "readmit_90d_C_TKA"
```

```
## [105] "readmit_date_C_TKA"
## [106] "ed_90d_C_TKA"
## [107] "ed_date_C_TKA"
## [108] "MUA_bi"
## [109] "C_MUA_bi"
## [110] "MUA_count_T"
## [111] "age_diff"
## [112] "los_total"
## [113] "op_time_total"
## [114] "date_diff"
## [115] "BTKA"
## [116] "Fac_age_C_TKA"
## [117] "Fac_age_TKA"
## [118] "redu_race"
## [119] "redu_tobacco"
## [120] "redu_tobacco_C_TKA"
## [121] "redu_ASA_C_TKA"
## [122] "redu_ASA"
## [123] "yeardiff"
```

```
sex <- MUA_data1[,c(1:4,6)]
sex_table <- table(sex[,4:5])

# fisher's exact test for count data
test <-chisq.test(sex_table)
```

```
## Warning in chisq.test(sex_table): Chi-squared approximation may be incorrect
```

```
fisher.test(sex_table)
```

```
##
## Fisher's Exact Test for Count Data
##
## data: sex_table
## p-value = 0.7526
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.1209991 3.2422035
## sample estimates:
## odds ratio
## 0.7317336
```

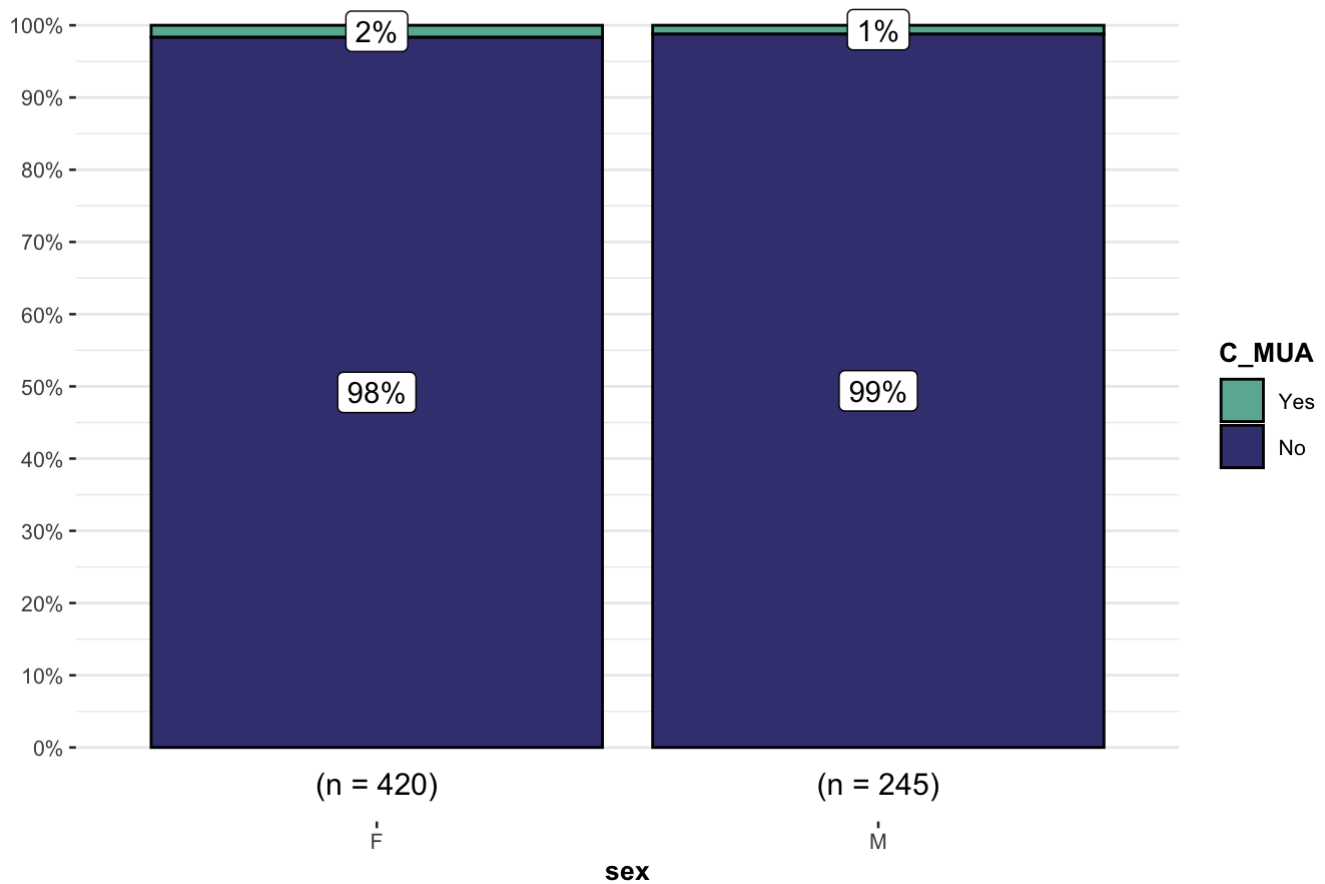
```
fisher.test(table(MUA_data1$C_MUA, MUA_data1$sex))
```

```
##  
## Fisher's Exact Test for Count Data  
##  
## data:  table(MUA_data1$C_MUA, MUA_data1$sex)  
## p-value = 0.7526  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
##  0.1209991 3.2422035  
## sample estimates:  
## odds ratio  
##  0.7317336
```

```
# combine plot and statistical test with ggbarstats  
library(ggstatsplot)  
ggbarstats(  
  MUA_data1, C_MUA, sex,  
  results.subtitle = FALSE,  
  subtitle = paste0(  
    "Fisher's exact test", ", p-value = ",  
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))  
  )  
) + scale_fill_manual(values=c("#69b3a2", "#404080"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which  
## will replace the existing scale.
```

Fisher's exact test, p-value = 0.903



b. fitting model

```
##### add sex variable - insignificant #####  
mua_glm1 = glm(C_MUA ~ sex, data= MUA_data1, family = "binomial")  
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ sex, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1833  -0.1833  -0.1833  -0.1570   2.9674
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.0775     0.3812  -10.70  <2e-16 ***
## sexM          -0.3128     0.6948   -0.45   0.653
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 103.58  on 663  degrees of freedom
## AIC: 107.58
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 <- glm(as.factor(Group2) ~ sex, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = as.factor(Group2) ~ sex, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5293   0.2887   0.3203   0.3203   0.3203
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.9444     0.2239  13.151  <2e-16 ***
## sexM          0.2126     0.3929   0.541   0.589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 250.61  on 664  degrees of freedom
## Residual deviance: 250.31  on 663  degrees of freedom
## AIC: 254.31
##
## Number of Fisher Scoring iterations: 6
```

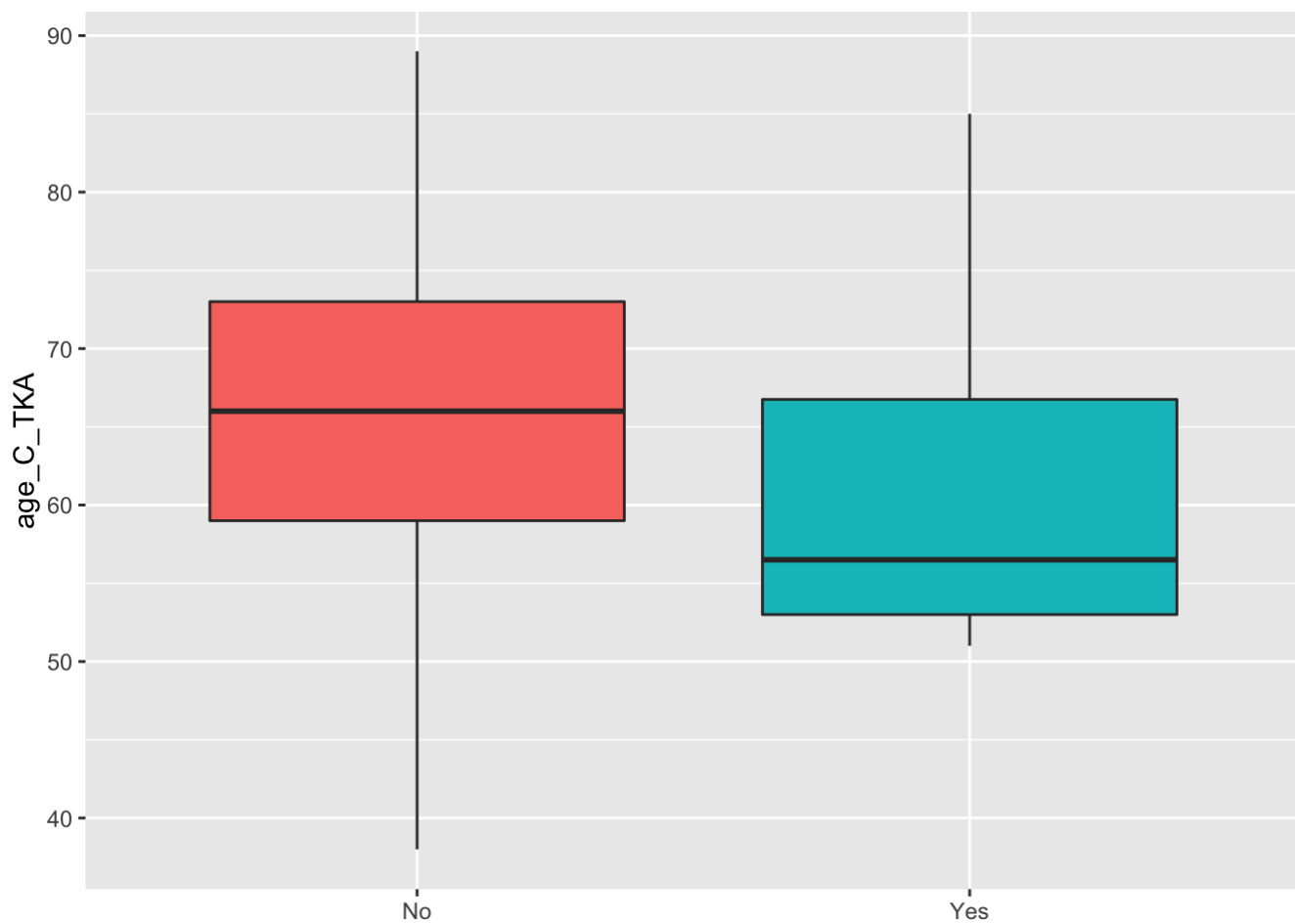
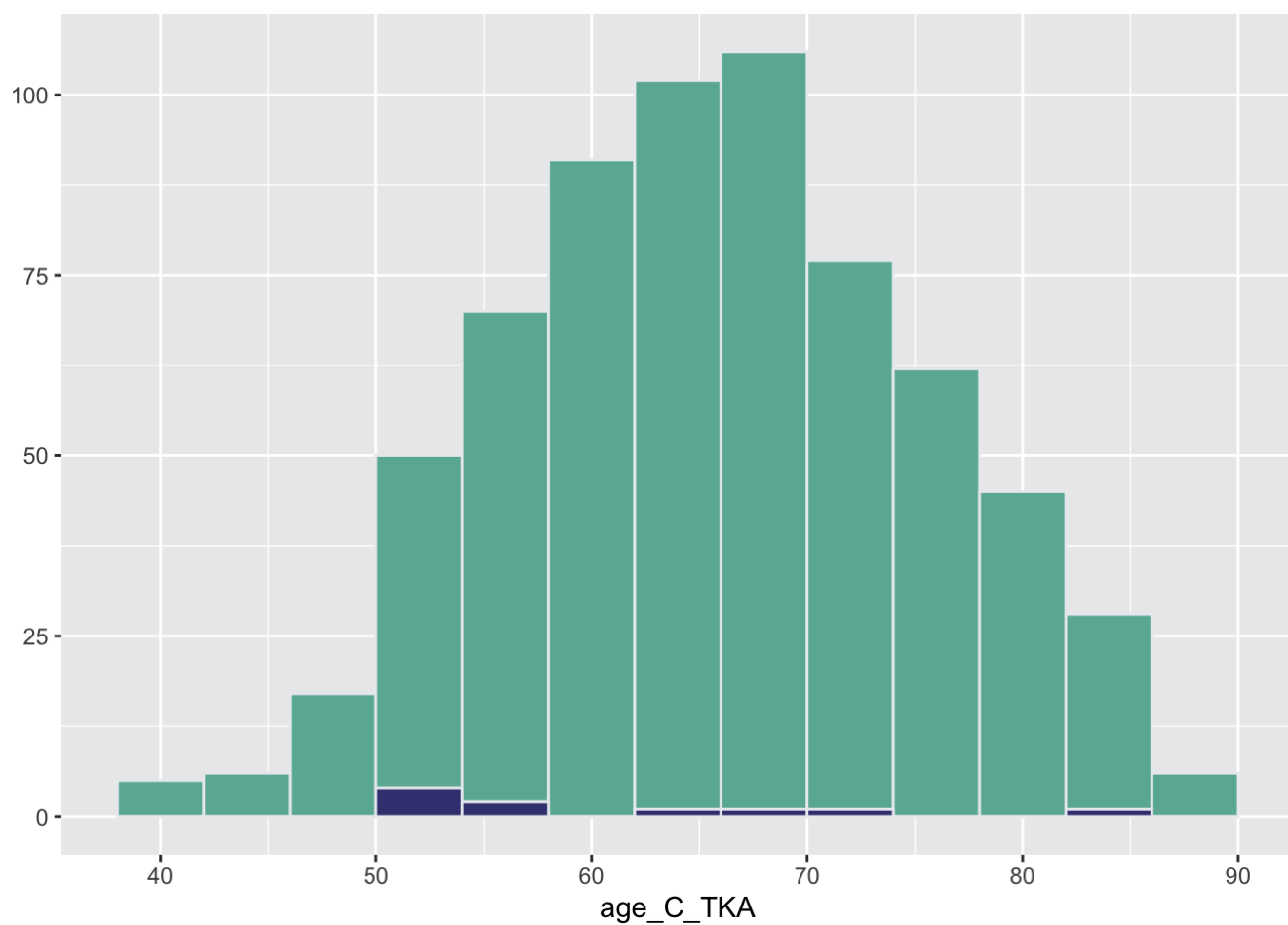
```
mua_glm1 = glm(C_MUA ~ MUA +sex, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + sex, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5321  -0.1541  -0.1541  -0.1376   3.0542
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.4276     0.4424 -10.008  < 2e-16 ***
## MUAYes        2.5440     0.7265   3.502 0.000463 ***
## sexM         -0.2271     0.7053  -0.322 0.747424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  95.142  on 662  degrees of freedom
## AIC: 101.14
##
## Number of Fisher Scoring iterations: 7
```

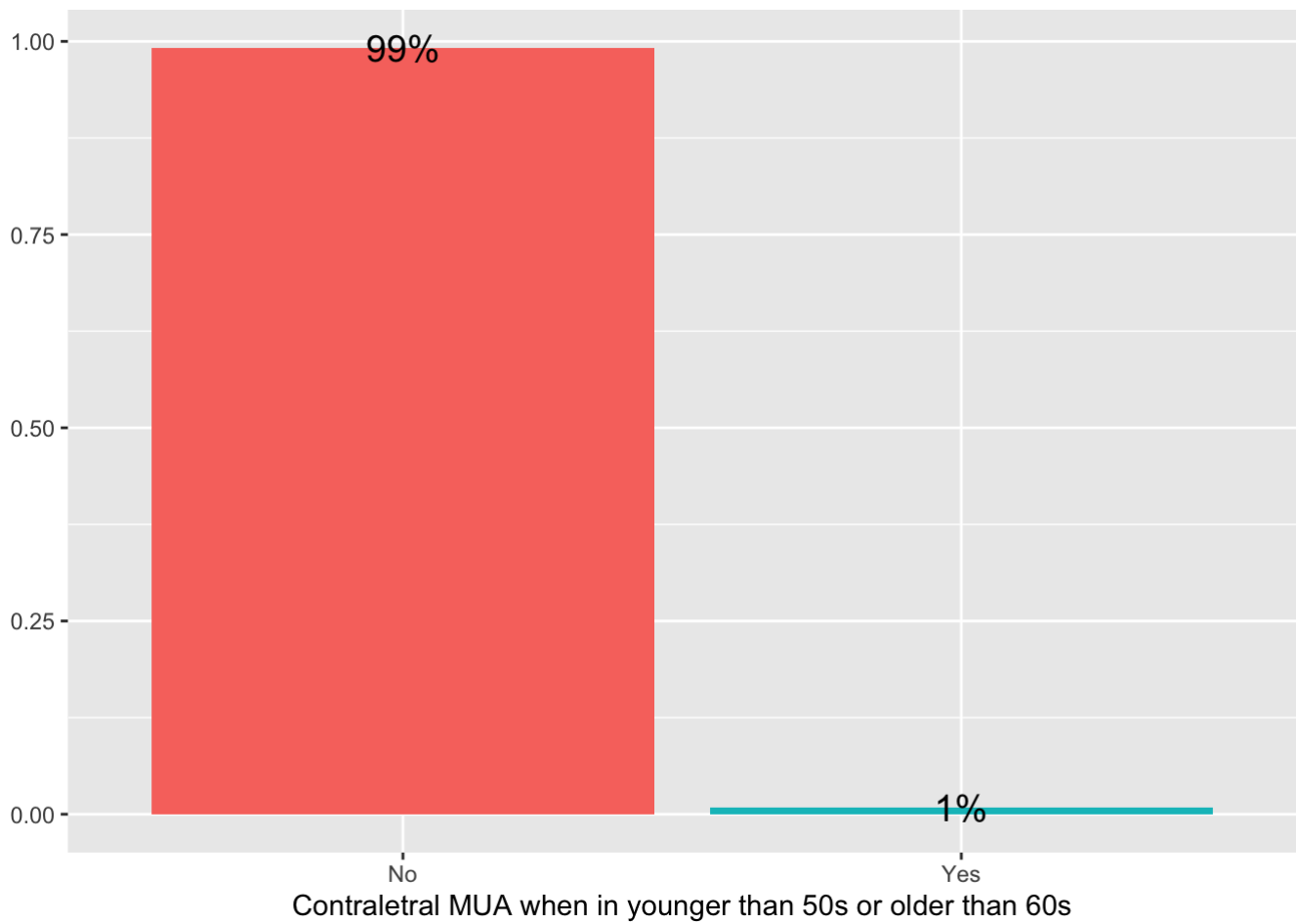
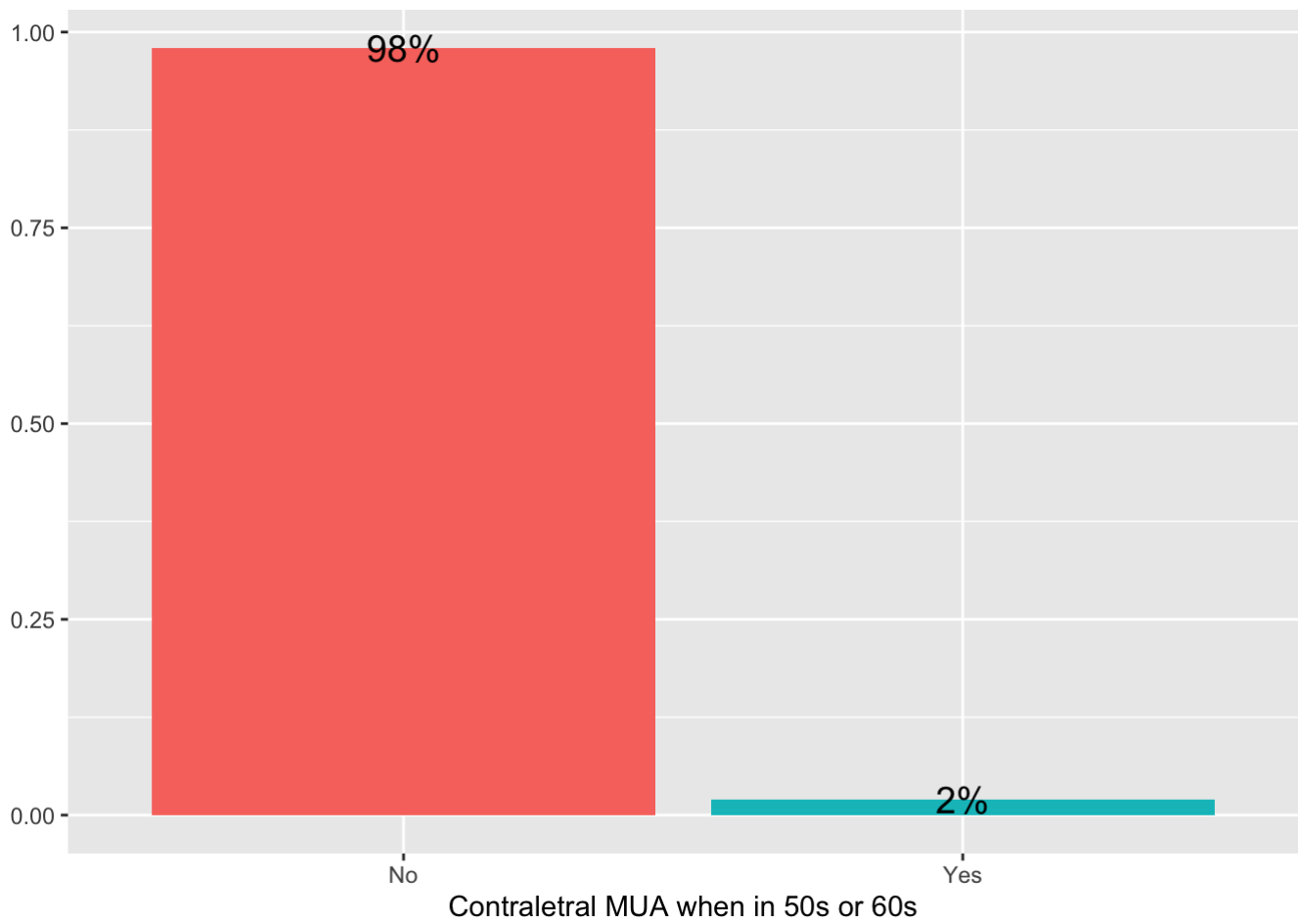
The sex variable is insignificant in these models.

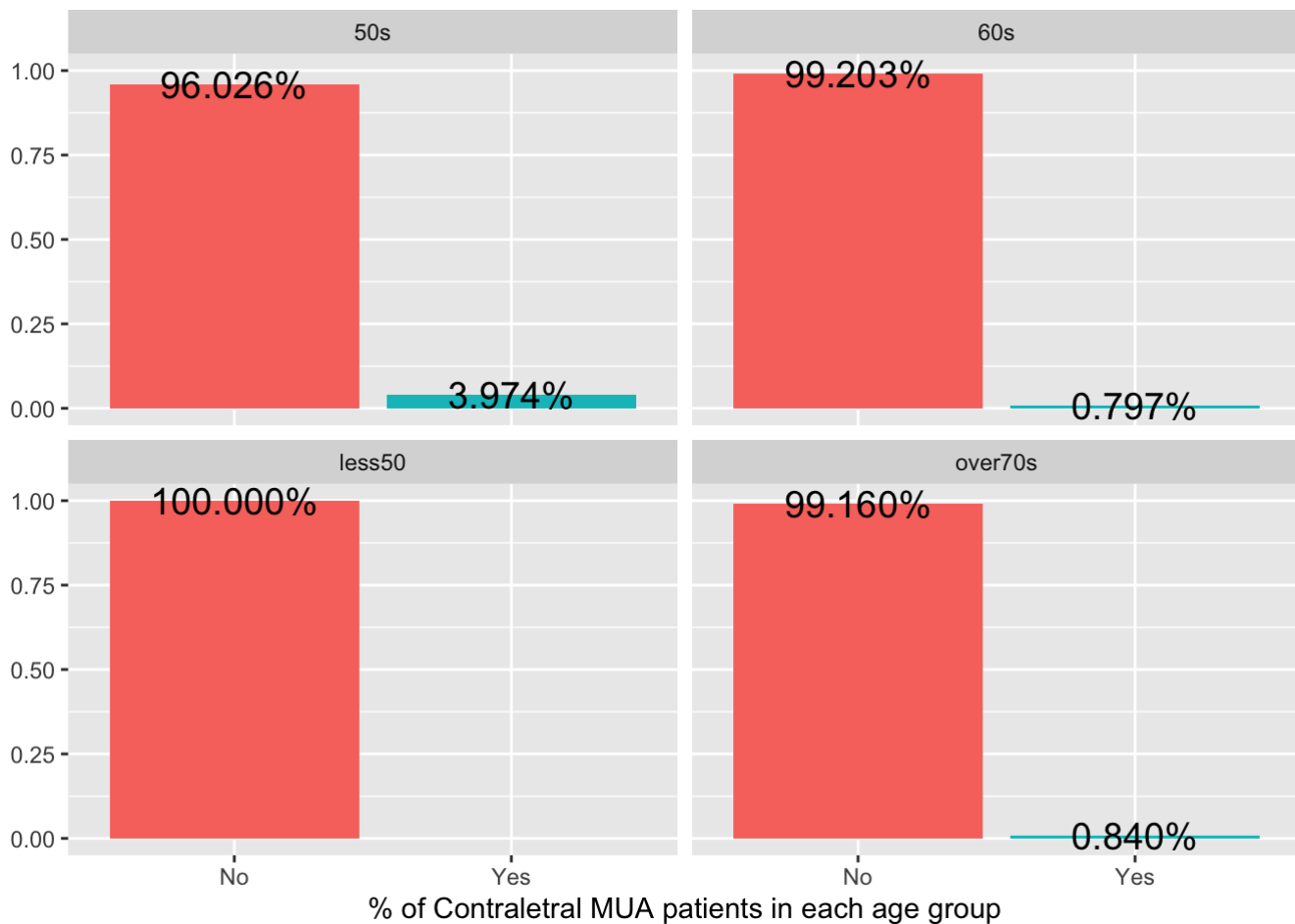
## (2) Age

### a. EDA









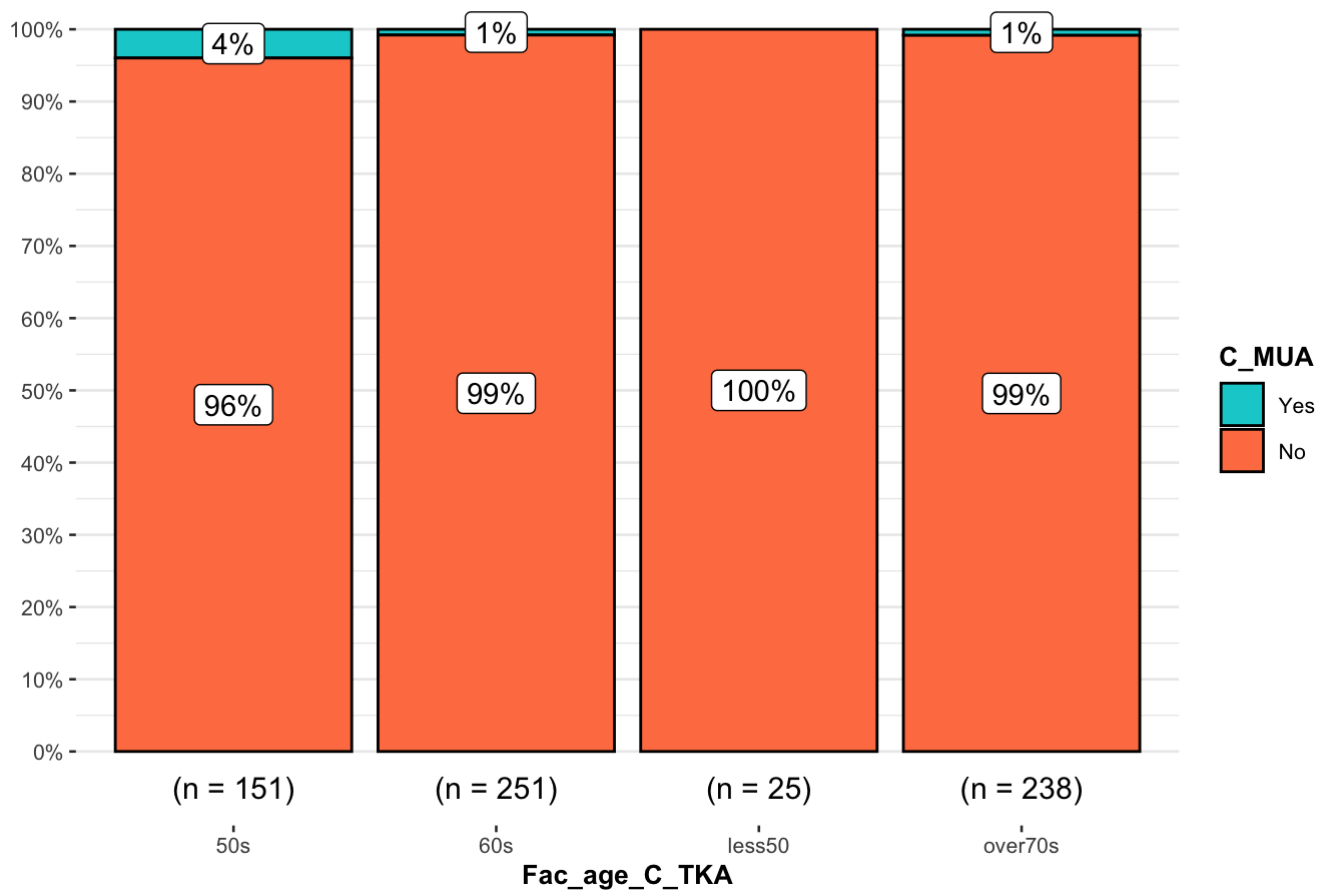
b.

```
# fisher's exact test for count data
test <- fisher.test(table(MUA_data1$C_MUA, MUA_data1$Fac_age_C_TKA))

# combine plot and statistical test with ggbarstats
library(ggstatsplot)
ggbarstats(
  MUA_data1, C_MUA, Fac_age_C_TKA,
  results.subtitle = FALSE,
  subtitle = paste0(
    "Fisher's exact test", ", p-value = ",
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))
  )
) + scale_fill_manual(values = c( "darkturquoise", "coral"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

Fisher's exact test, p-value = 0.081



c. regression model

i. with numerical age variable

```
mua_glm1 = glm(C_MUA ~ age_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ age_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3588  -0.1982  -0.1623  -0.1328   3.2929
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.52301     2.07238  -0.252   0.8008
## age_C_TKA   -0.05758     0.03357  -1.715   0.0863 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 100.79  on 663  degrees of freedom
## AIC: 104.79
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA + age_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1) #exp(-0.04859) = 0.9525716 : as age increase 1, the odds of getting C
_MUA decrease 5%..? Not true
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + age_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6582  -0.1677  -0.1416  -0.1195   3.3242
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.39109     2.28043  -0.610   0.54185
## MUAYes       2.35246     0.73571   3.198   0.00139 **
## age_C_TKA   -0.04859     0.03633  -1.337   0.18108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  93.434  on 662  degrees of freedom
## AIC: 99.434
##
## Number of Fisher Scoring iterations: 7
```

*# numeric age variable seems insignificant and not to make sense. EDA shows that when the patients are in 50s, 60s, the odds of getting MUA seems higher*

ii. with factorized age variable

```
MUA_data1$Fac_age_C_TKA <- ifelse(MUA_data1$age_C_TKA<50,"less50",  
                                  ifelse(MUA_data1$age_C_TKA<60, "50s",  
                                           ifelse(MUA_data1$age_C_TKA<70, "60s","over70s")))  
  
MUA_data1$Fac_age_C_TKA2 <- ifelse(MUA_data1$Fac_age_C_TKA %in% c("50s","60s"),"50/60",  
                                   "younger than 50|older than 60")  
  
MUA_data1 %>% group_by(Fac_age_C_TKA2) %>% count(Group2)
```

```
## # A tibble: 4 × 3  
## # Groups:   Fac_age_C_TKA2 [2]  
##   Fac_age_C_TKA2      Group2      n  
##   <chr>             <chr> <int>  
## 1 50/60             MUA      25  
## 2 50/60             No_MUA    377  
## 3 younger than 50|older than 60 MUA      6  
## 4 younger than 50|older than 60 No_MUA    257
```

```
mua_glm1 = glm(C_MUA ~ Fac_age_C_TKA2, data= MUA_data1, family = "binomial")  
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ Fac_age_C_TKA2, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2005  -0.2005  -0.2005  -0.1236   3.1238
##
## Coefficients:
##                                Estimate Std. Error z value
## (Intercept)                   -3.8969     0.3571 -10.912
## Fac_age_C_TKA2younger than 50|older than 60 -0.9745     0.7945  -1.226
##                                Pr(>|z|)
## (Intercept)                   <2e-16 ***
## Fac_age_C_TKA2younger than 50|older than 60     0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 102.01  on 663  degrees of freedom
## AIC: 106.01
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA +Fac_age_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + Fac_age_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7063  -0.1234  -0.1234  -0.1089   3.2032
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.5885     0.4926  -7.284 3.24e-13 ***
## MUAYes         2.3271     0.7567   3.075  0.0021 **
## Fac_age_C_TKA60s -1.5360     0.8345  -1.841  0.0657 .
## Fac_age_C_TKAless50 -15.2348  1267.5272  -0.012  0.9904
## Fac_age_C_TKAover70s -1.2857     0.8478  -1.516  0.1294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  89.844  on 660  degrees of freedom
## AIC: 99.844
##
## Number of Fisher Scoring iterations: 17
```

```
mua_glm1 = glm(C_MUA ~ (MUA +Fac_age_C_TKA2)^2, data= MUA_data1, family = "binomial")
summary(mua_glm1) # interaction term is insignificant
```

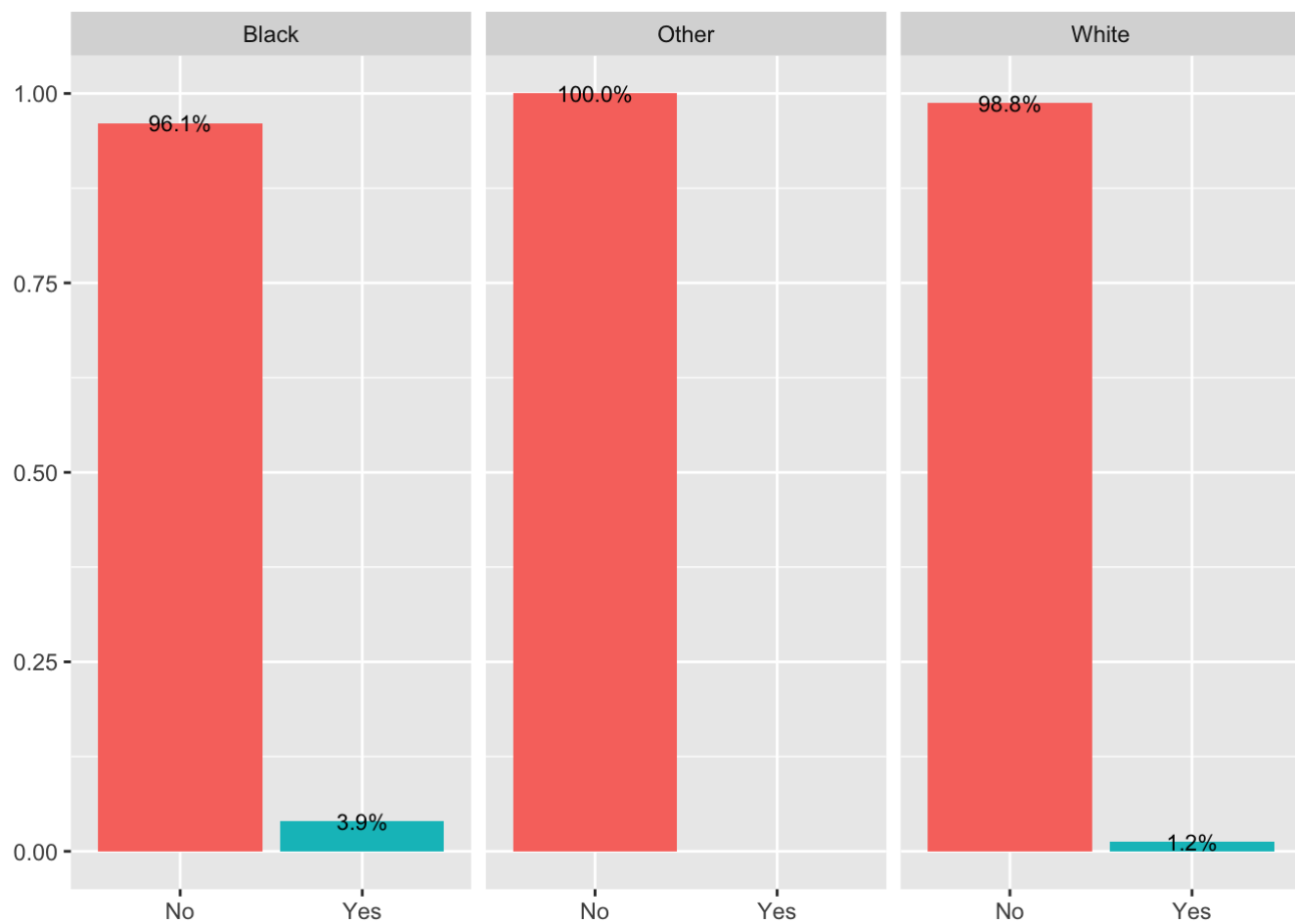
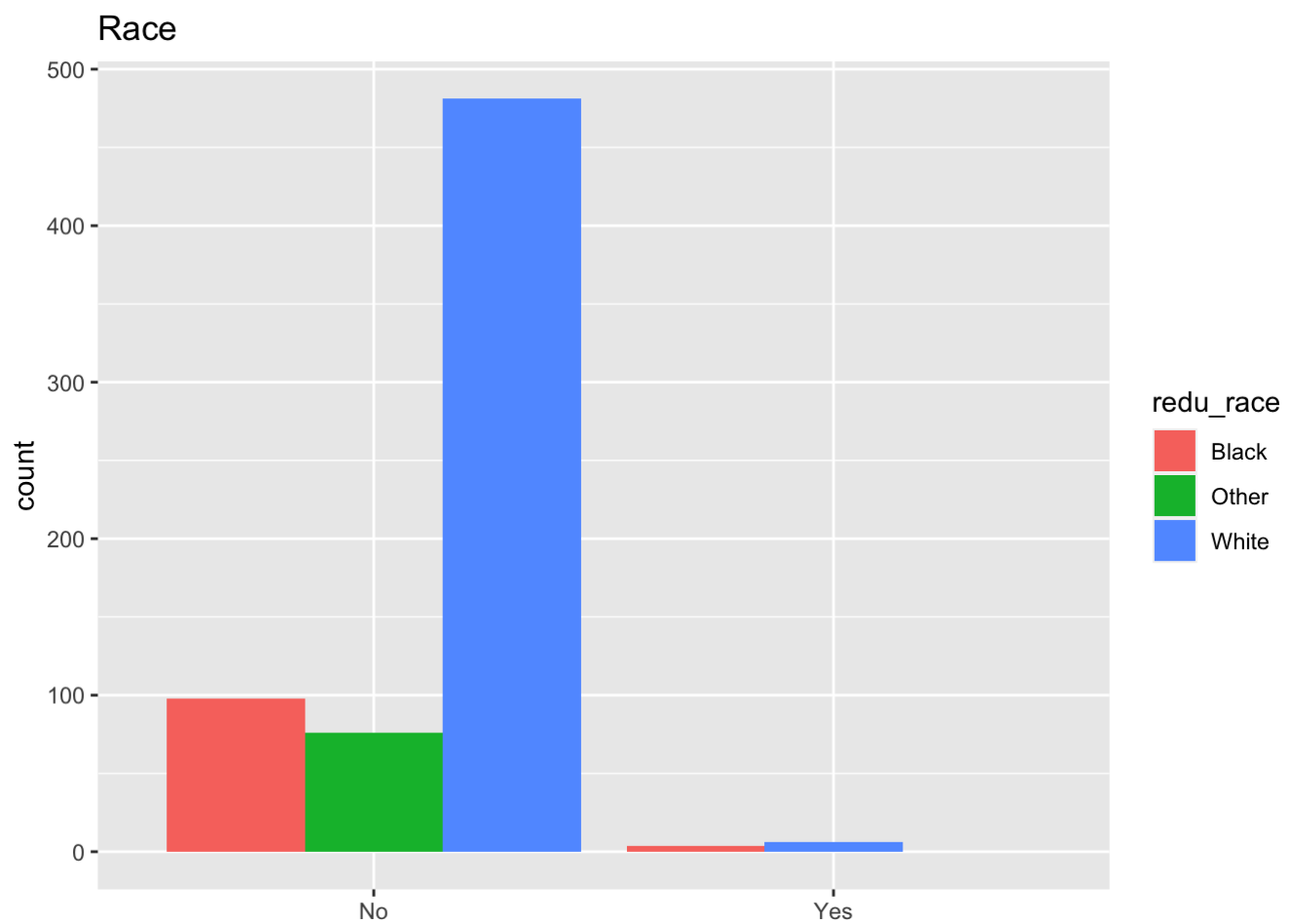
```
##
## Call:
## glm(formula = C_MUA ~ (MUA + Fac_age_C_TKA2)^2, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -0.5701   -0.1623   -0.1623   -0.1245    3.1189
##
## Coefficients:
##                                     Estimate Std. Error z value
## (Intercept)                      -4.3228     0.4502  -9.603
## MUAYes                           2.5882     0.7712   3.356
## Fac_age_C_TKA2younger than 50|older than 60  -0.5331     0.8406  -0.634
## MUAYes:Fac_age_C_TKA2younger than 50|older than 60 -14.2983  1199.7728  -0.012
##                                     Pr(>|z|)
## (Intercept)                      < 2e-16 ***
## MUAYes                           0.000791 ***
## Fac_age_C_TKA2younger than 50|older than 60  0.525921
## MUAYes:Fac_age_C_TKA2younger than 50|older than 60 0.990491
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  93.642  on 661  degrees of freedom
## AIC: 101.64
##
## Number of Fisher Scoring iterations: 15
```

- when the patients are in 50s or 60s, age is significant. interaction is insignificant
- keep age in the model.

### (3) Race

#### a. EDA



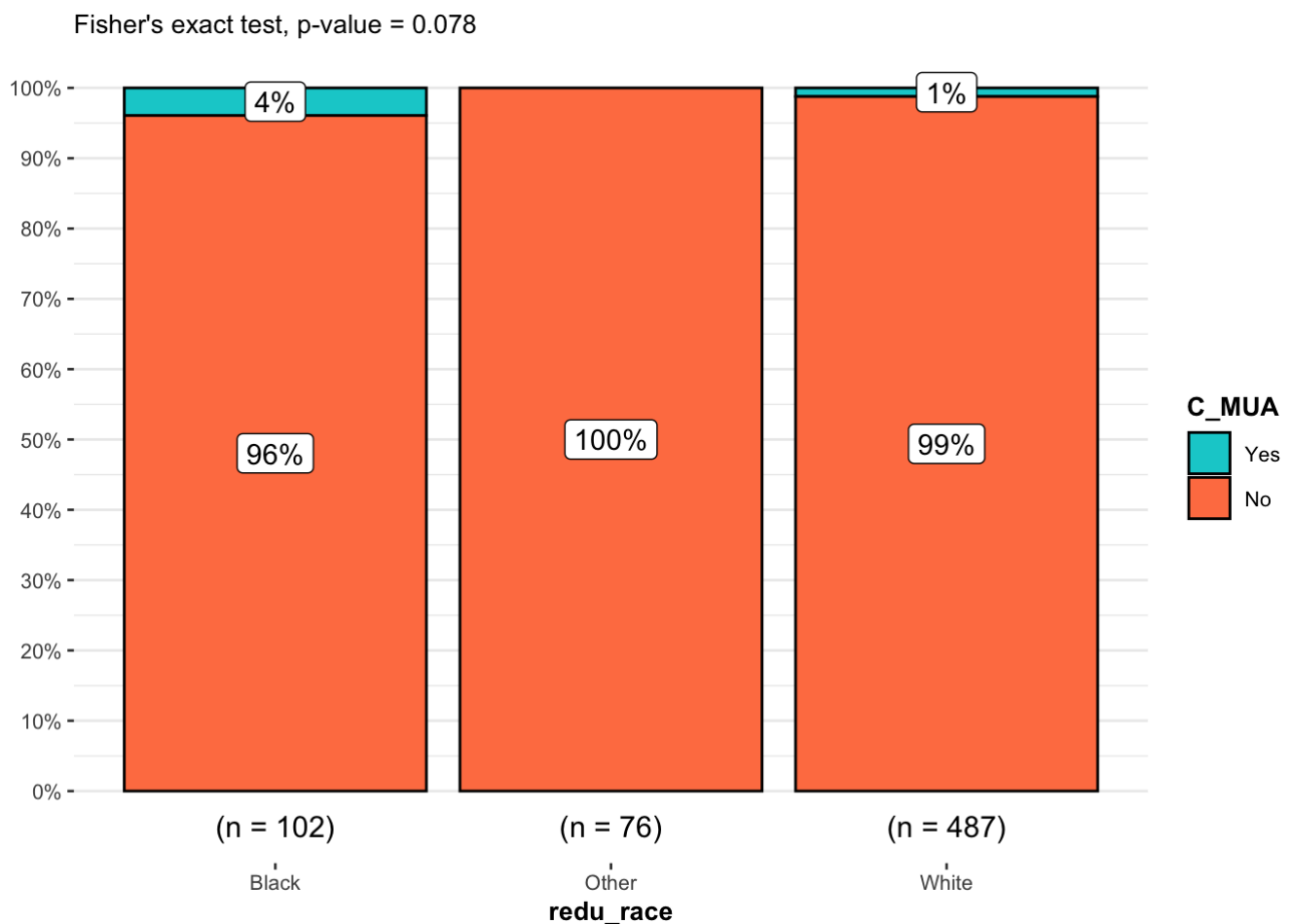


## b. chi-squared test

```
# fisher's exact test for count data
test<- fisher.test(table(MUA_data1$C_MUA, MUA_data1$redu_race))

# combine plot and statistical test with ggbarstats
ggbarstats(
  MUA_data1, C_MUA, redu_race,
  results.subtitle = FALSE,
  subtitle = paste0(
    "Fisher's exact test", ", ", p-value = ",
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))
  )
) + scale_fill_manual(values = c( "darkturquoise", "coral"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```



## b. fitting model

```
##### race variable
# the stand.error is way tooooooooooo big..
# race black is significant..?

mua_glm1 <- glm(C_MUA ~ as.factor(redu_race), data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ as.factor(redu_race), family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2829  -0.1575  -0.1575  -0.1575   2.9653
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -3.1987     0.5101  -6.271  3.6e-10 ***
## as.factor(redu_race)Other -16.3674  1233.5700  -0.013   0.9894
## as.factor(redu_race)White  -1.1854     0.6549  -1.810   0.0703 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  98.434  on 662  degrees of freedom
## AIC: 104.43
##
## Number of Fisher Scoring iterations: 18
```

```
mua_glm1 = glm(C_MUA ~ MUA+ as.factor(redu_race), data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + as.factor(redu_race), family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7192  -0.1397  -0.1397  -0.1397   3.0444
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -3.6508     0.5880  -6.209 5.33e-10 ***
## MUAYes           2.4307     0.7461   3.258 0.00112 **
## as.factor(redu_race)Other -16.1944  1194.7891  -0.014 0.98919
## as.factor(redu_race)White  -0.9734     0.6787  -1.434 0.15151
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  90.802  on 661  degrees of freedom
## AIC: 98.802
##
## Number of Fisher Scoring iterations: 18
```

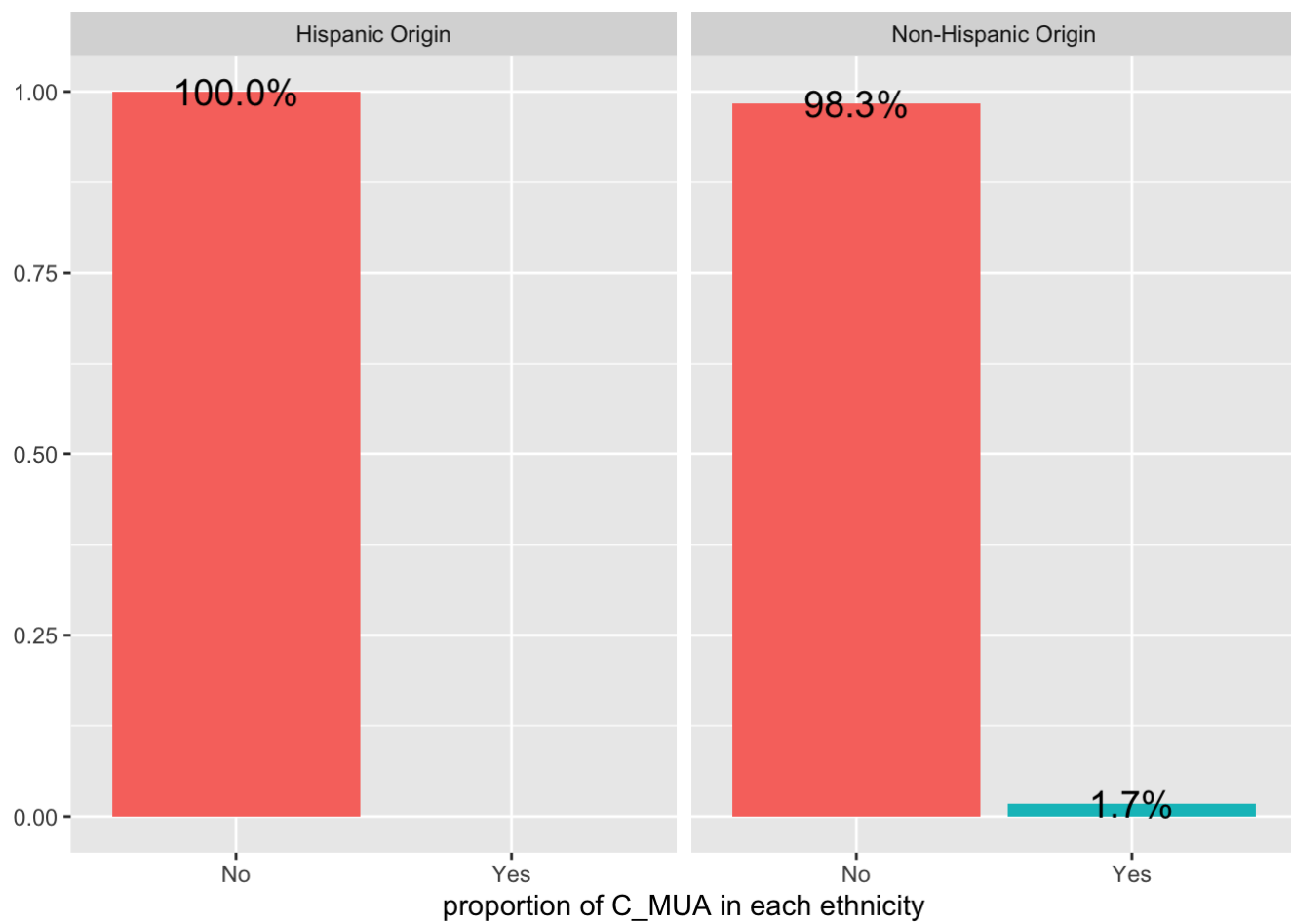
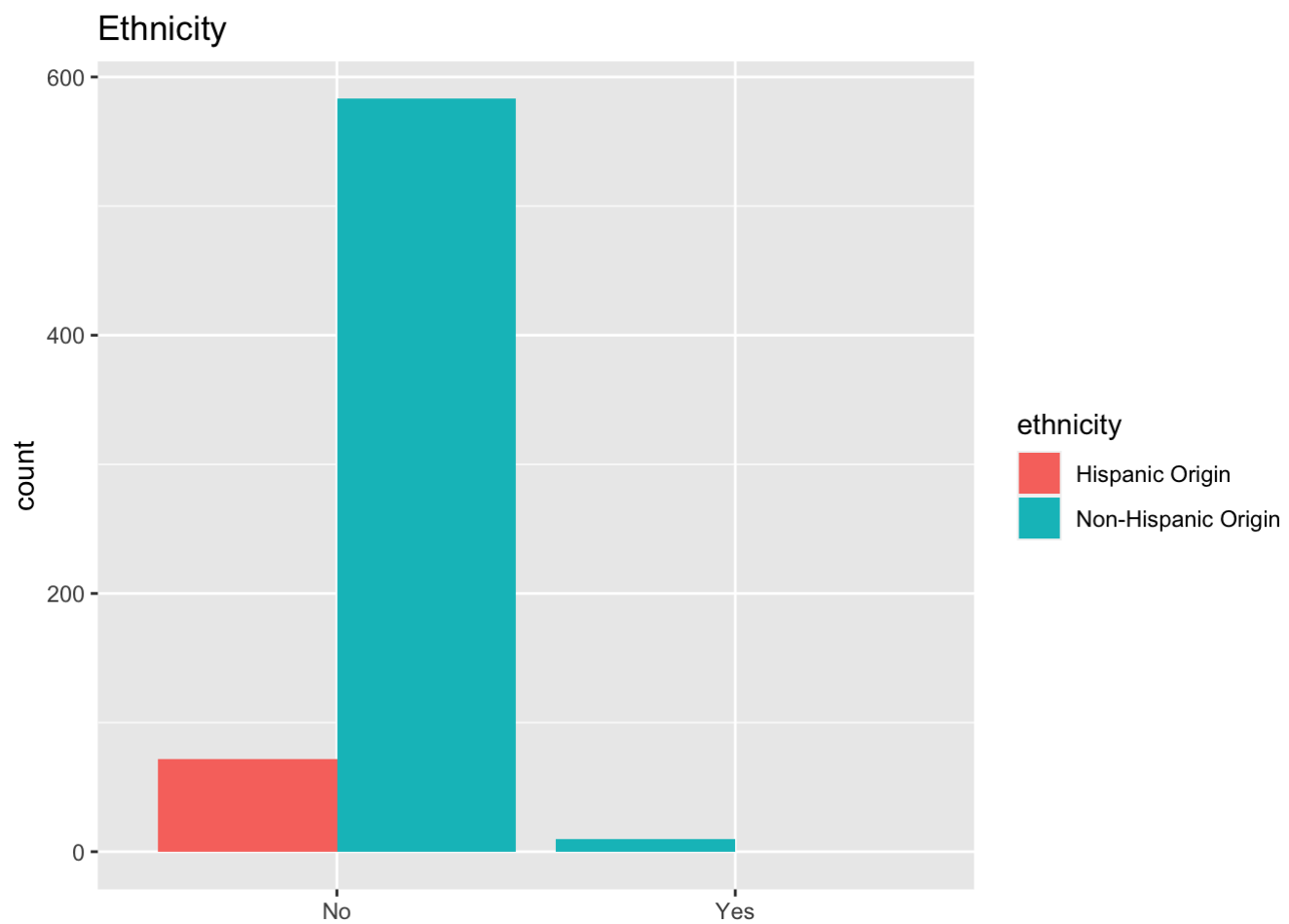
```
mua_glm1 = glm(C_MUA ~ (MUA+ as.factor(redu_race))^2, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ (MUA + as.factor(redu_race))^2, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8203  -0.1458  -0.1458  -0.1458   3.0165
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -3.8395     0.7147  -5.372 7.77e-08 ***
## MUAYes                     2.9232     1.1003   2.657  0.00789 **
## as.factor(redu_race)Other  -15.7266  1258.6622  -0.012  0.99003
## as.factor(redu_race)White   -0.6996     0.8443  -0.829  0.40735
## MUAYes:as.factor(redu_race)Other -2.9232  6335.1266   0.000  0.99963
## MUAYes:as.factor(redu_race)White  -0.9491     1.5779  -0.601  0.54752
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  90.425  on 659  degrees of freedom
## AIC: 102.42
##
## Number of Fisher Scoring iterations: 18
```

```
#mua_glm1 = glm(C_MUA ~ MUA+ as.factor(redu_race)+Fac_age_C_TKA, data= #MUA_data1, famil
y = "binomial")
#summary(mua_glm1)
```

#### (4). Ethnicity

##### a. EDA

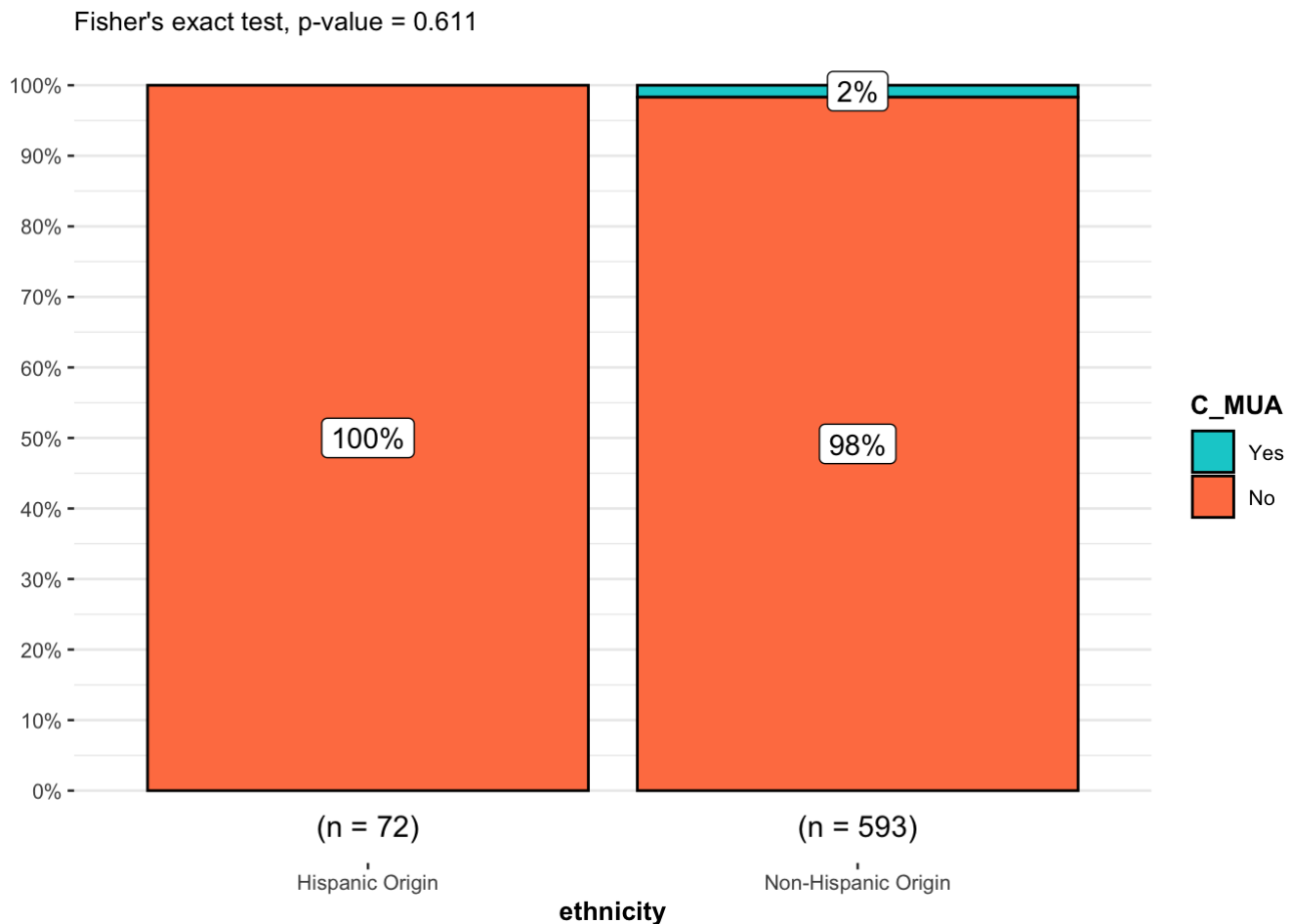


## b. chi-squared test

```
# fisher's exact test for count data
test<- fisher.test(table(MUA_data1$C_MUA, MUA_data1$ethnicity))

# combine plot and statistical test with ggbarstats
ggbarstats(
  MUA_data1,C_MUA,ethnicity,
  results.subtitle = FALSE,
  subtitle = paste0(
    "Fisher's exact test", ", ", p-value = ",
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))
  )
) + scale_fill_manual(values = c( "darkturquoise","coral"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```



## C. regression model

```
##### ethnicity variable
mua_glm1 = glm(C_MUA ~ ethnicity, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ ethnicity, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1844  -0.1844  -0.1844  -0.1844   2.8575
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -19.57    1267.37  -0.015    0.988
## ethnicityNon-Hispanic Origin    15.50    1267.37   0.012    0.990
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 101.48  on 663  degrees of freedom
## AIC: 105.48
##
## Number of Fisher Scoring iterations: 18
```

```
mua_glm1 = glm(C_MUA ~ MUA+ethnicity, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

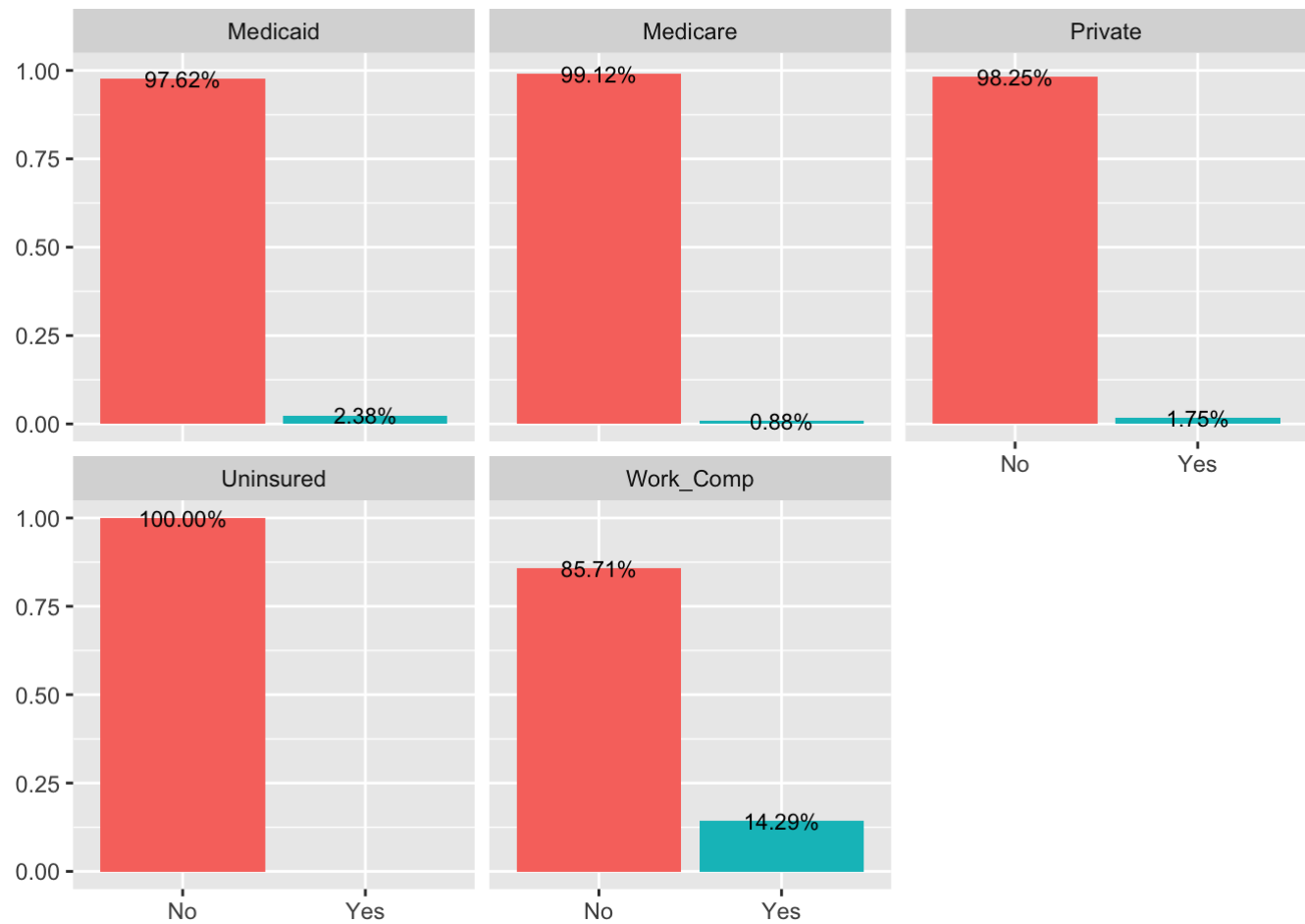
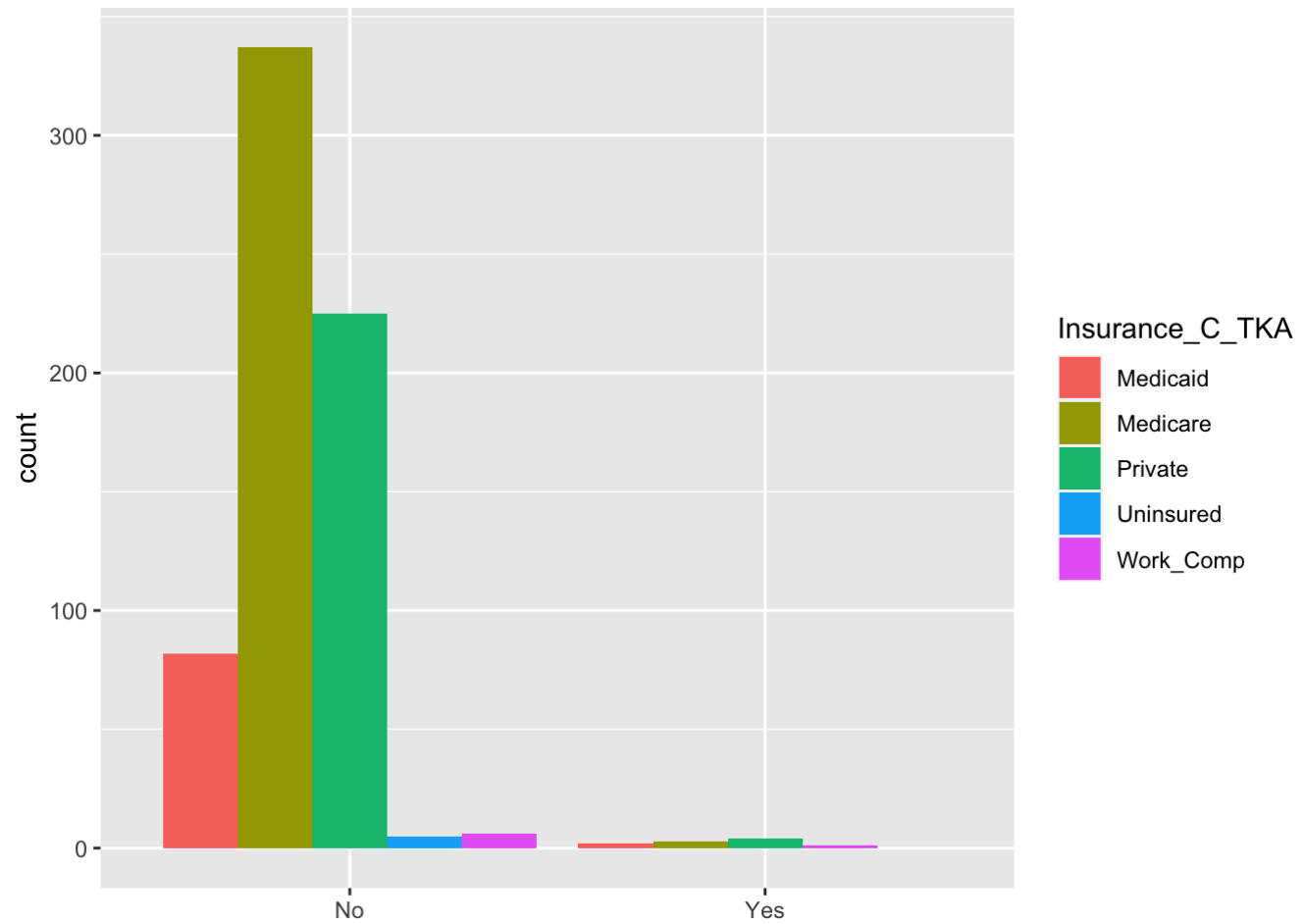
```
##
## Call:
## glm(formula = C_MUA ~ MUA + ethnicity, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5863  -0.1567  -0.1567  -0.1567   2.9687
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -20.1270    1190.7128  -0.017  0.986514
## MUAYes              2.7205      0.7352   3.701 0.000215 ***
## ethnicityNon-Hispanic Origin    15.7326    1190.7128   0.013 0.989458
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  92.183  on 662  degrees of freedom
## AIC: 98.183
##
## Number of Fisher Scoring iterations: 18
```

```
# the stand.error is tooooooooooo big..
```



**(5) Insurance**

a.EDA

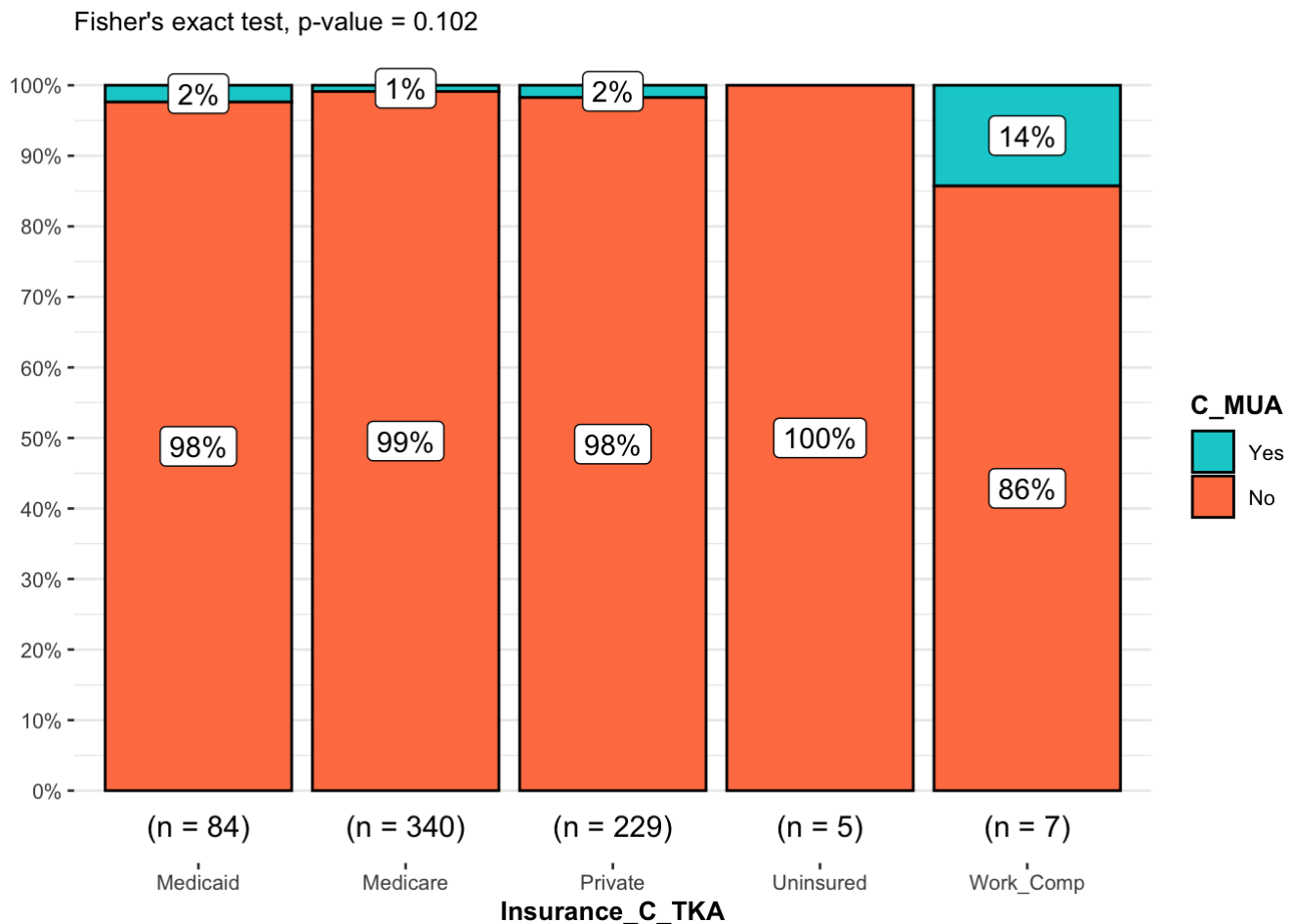


## b. chi-squared test

```
# fisher's exact test for count data
test<- fisher.test(table(MUA_data1$C_MUA, MUA_data1$Insurance_C_TKA))

# combine plot and statistical test with ggbarstats
ggbarstats(
  MUA_data1,C_MUA,Insurance_C_TKA,
  results.subtitle = FALSE,
  subtitle = paste0(
    "Fisher's exact test", ", ", p-value = ",
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))
  )
) + scale_fill_manual(values = c( "darkturquoise","coral"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```



## c. Regression model

```
mua_glm1 = glm(C_MUA ~ Insurance_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ Insurance_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5553  -0.1877  -0.1331  -0.1331   3.0758
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.7136     0.7157  -5.189 2.12e-07 ***
## Insurance_C_TKAMedicare    -1.0079     0.9211  -1.094   0.274
## Insurance_C_TKAPrivate     -0.3162     0.8756  -0.361   0.718
## Insurance_C_TKAUninsured  -13.8525    1769.2578  -0.008   0.994
## Insurance_C_TKAWork_Comp     1.9218     1.2957   1.483   0.138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  99.309  on 660  degrees of freedom
## AIC: 109.31
##
## Number of Fisher Scoring iterations: 16
```

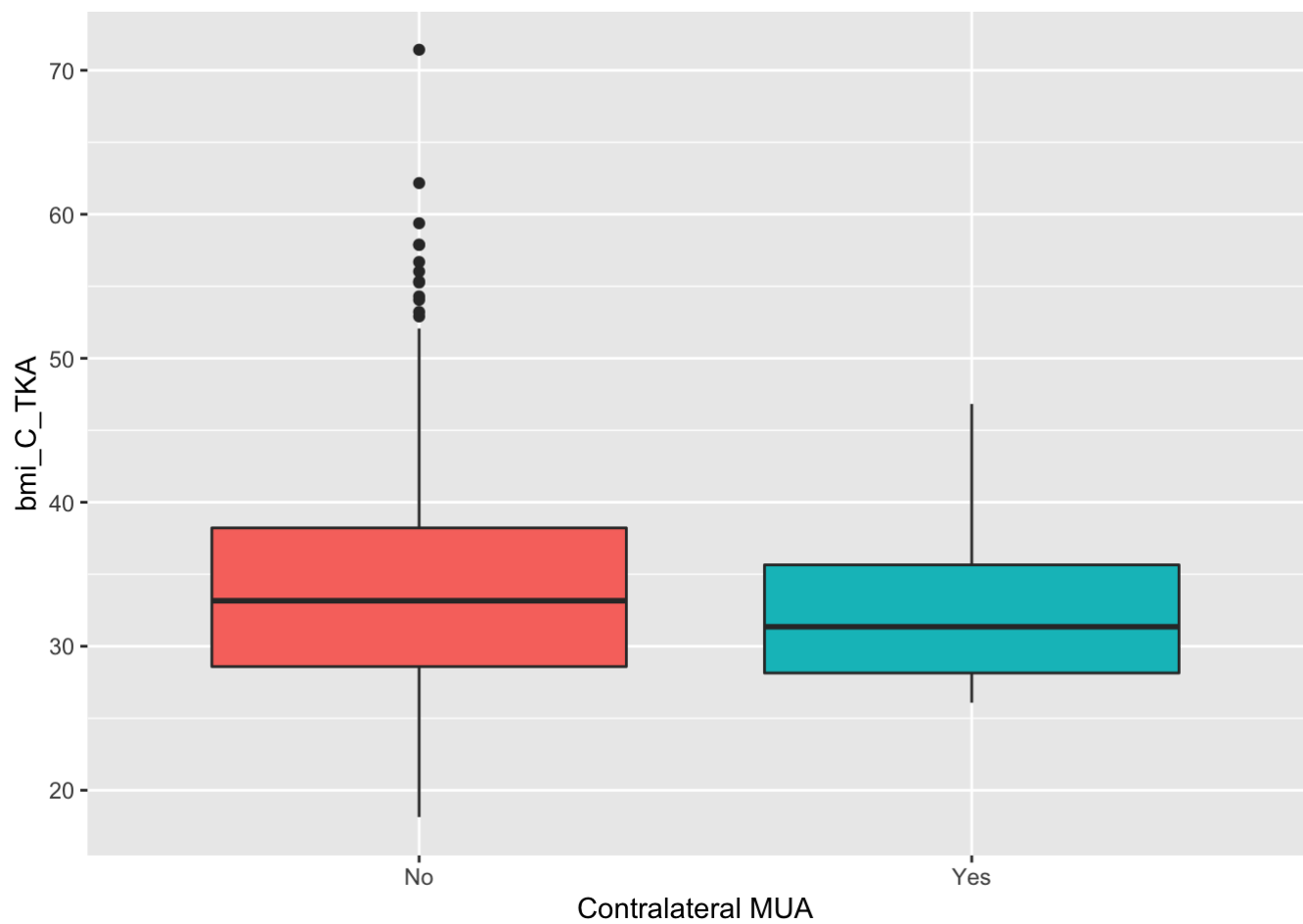
```
mua_glm1 = glm(C_MUA ~ MUA+Insurance_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + Insurance_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5375  -0.1554  -0.1270  -0.1270   3.1061
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.1649     0.7848  -5.307 1.11e-07 ***
## MUAYes         2.3032     0.7680   2.999  0.00271 **
## Insurance_C_TKAMedicare -0.6509     0.9582  -0.679  0.49693
## Insurance_C_TKAPrivate  -0.2461     0.8928  -0.276  0.78281
## Insurance_C_TKAUninsured -13.4012  1769.2578  -0.008  0.99396
## Insurance_C_TKAWork_Comp  1.8001     1.3862   1.299  0.19408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  92.535  on 659  degrees of freedom
## AIC: 104.53
##
## Number of Fisher Scoring iterations: 16
```

## (6) BMI

### a. EDA

Variable	N	No, N = 655 <sup>1</sup>	Yes, N = 10 <sup>1</sup>
<b>bmi_C_TKA</b>	665	33 (29, 38)	31 (28, 36)
<sup>1</sup> Median (IQR)			



b. Regression model

```
mua_glm1 = glm(C_MUA ~ bmi_C_TKA, data= MUA_data1, family = "binomial")  
summary(mua_glm1)
```

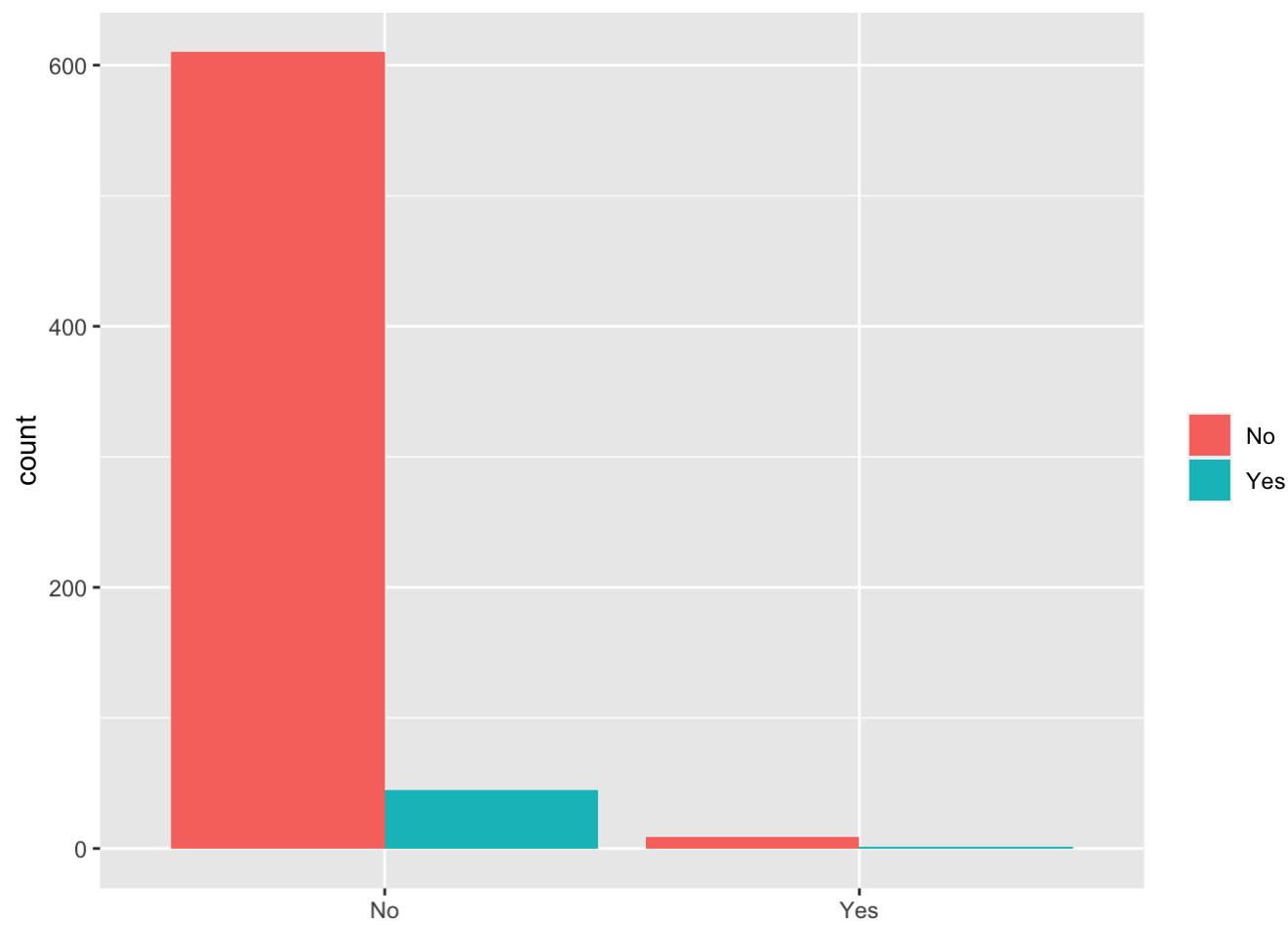
```
##
## Call:
## glm(formula = C_MUA ~ bmi_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2113  -0.1849  -0.1747  -0.1629   3.0114
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.32716     1.55742  -2.136   0.0327 *
## bmi_C_TKA   -0.02555     0.04641  -0.550   0.5820
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 103.47  on 663  degrees of freedom
## AIC: 107.47
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA+bmi_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + bmi_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5869  -0.1619  -0.1492  -0.1351   3.1608
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.33262     1.71153  -1.947 0.051516 .
## MUAYes       2.60529     0.73145   3.562 0.000368 ***
## bmi_C_TKA   -0.03536     0.05159  -0.685 0.493086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  94.744  on 662  degrees of freedom
## AIC: 100.74
##
## Number of Fisher Scoring iterations: 7
```

(7) tobacco

a. EDA



b. Regression Model

```
mua_glm1 = glm(C_MUA ~ redu_tobacco_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```



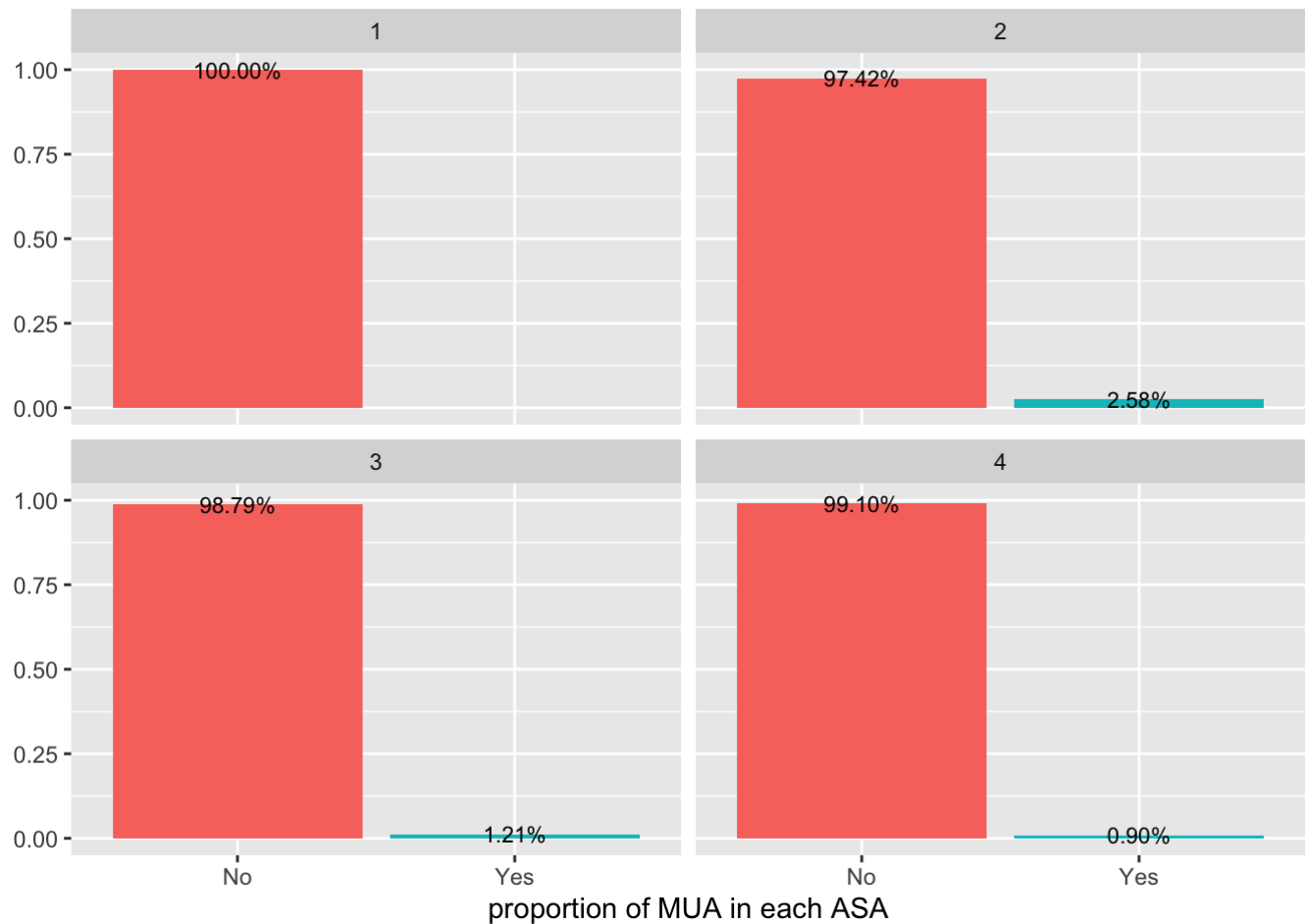
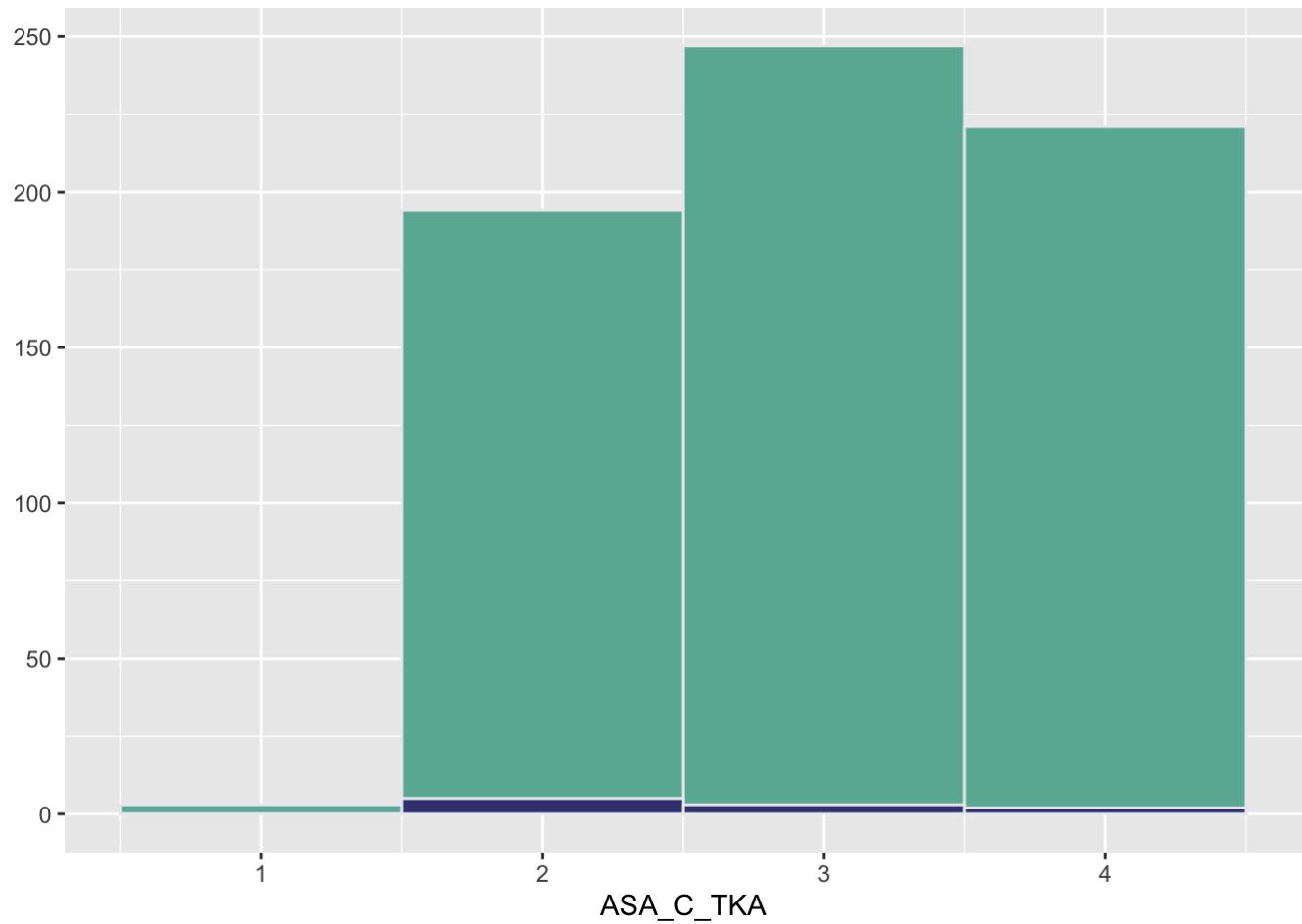
```
##
## Call:
## glm(formula = C_MUA ~ redu_tobacco_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2097  -0.1711  -0.1711  -0.1711   2.9089
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -4.2162     0.3358 -12.556  <2e-16 ***
## redu_tobacco_C_TKAYes  0.4096     1.0654   0.384   0.701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 103.66  on 663  degrees of freedom
## AIC: 107.66
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA+redu_tobacco_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + redu_tobacco_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5168  -0.1424  -0.1424  -0.1424   3.0321
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.5867     0.4103 -11.178 < 2e-16 ***
## MUAYes         2.6408     0.7412   3.563 0.000367 ***
## redu_tobacco_C_TKAYes 0.7800     1.0911   0.715 0.474700
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  94.821  on 662  degrees of freedom
## AIC: 100.82
##
## Number of Fisher Scoring iterations: 7
```

## (8) ASA

a. EDA

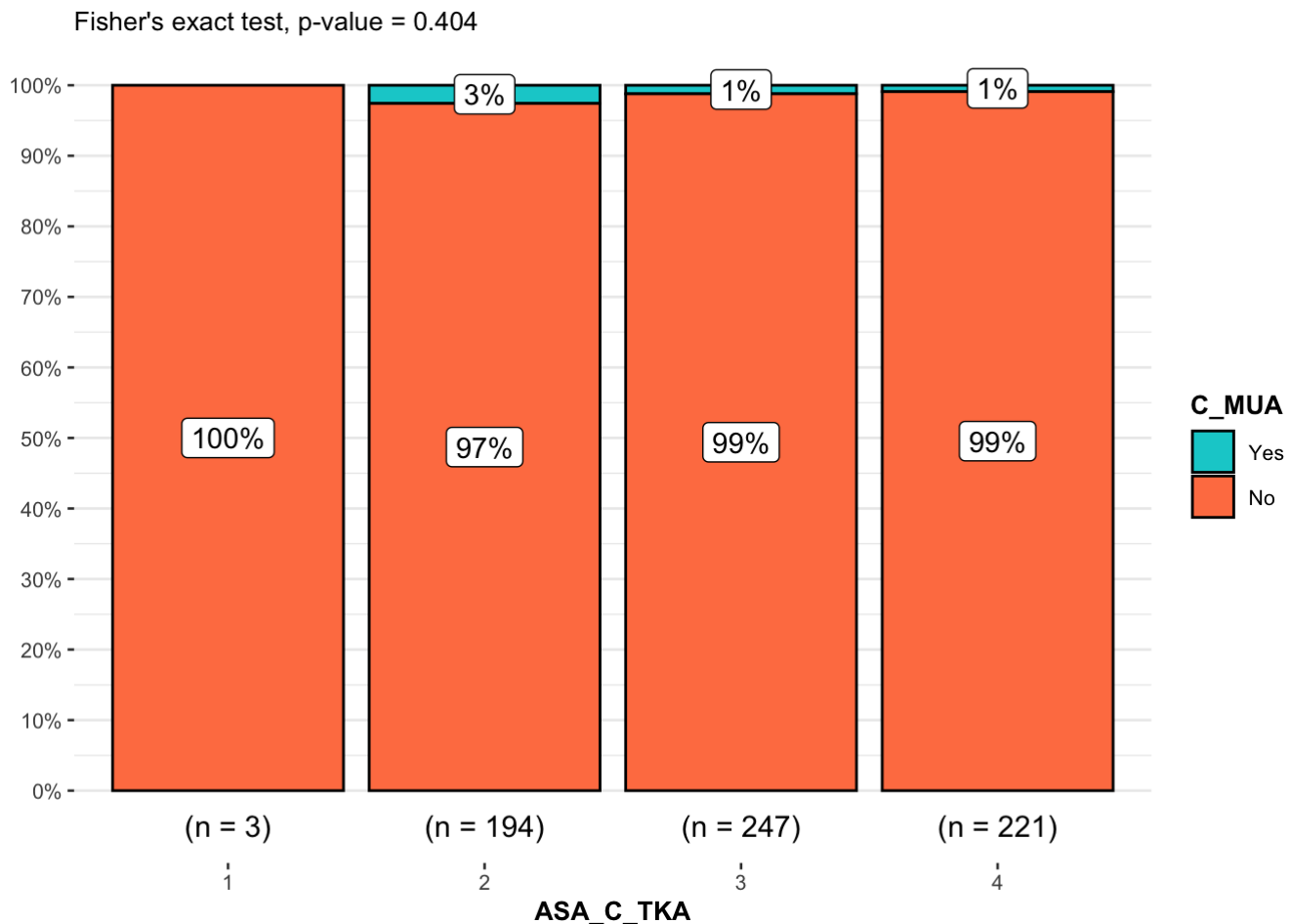


## b. fisher

```
# fisher's exact test for count data
test<- fisher.test(table(MUA_data1$C_MUA, MUA_data1$ASA_C_TKA))

# combine plot and statistical test with ggbarstats
ggbarstats(
  MUA_data1,C_MUA,ASA_C_TKA,
  results.subtitle = FALSE,
  subtitle = paste0(
    "Fisher's exact test", ", p-value = ",
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))
  )
) + scale_fill_manual(values = c( "darkturquoise","coral"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```



## b. Regression model

```
mua_glm1 = glm(C_MUA ~ as.factor(ASA_C_TKA), data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ as.factor(ASA_C_TKA), family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.2285   -0.2285   -0.1563   -0.1348    3.0676
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -16.57    1385.38  -0.012    0.990
## as.factor(ASA_C_TKA)2    12.93    1385.38   0.009    0.993
## as.factor(ASA_C_TKA)3    12.17    1385.38   0.009    0.993
## as.factor(ASA_C_TKA)4    11.87    1385.38   0.009    0.993
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 101.68  on 661  degrees of freedom
## AIC: 109.68
##
## Number of Fisher Scoring iterations: 15
```

```
# standard error is too big so reduced the factors into binary
```

```
MUA_data1$redu_ASA_C_TKA <- ifelse(MUA_data1$ASA_C_TKA %in% c(2,3),"2/3","else")
```

```
mua_glm1 = glm(C_MUA ~ as.factor(redu_ASA_C_TKA), data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ as.factor(redu_ASA_C_TKA), family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1913  -0.1913  -0.1913  -0.1339   3.0720
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -3.9913     0.3568 -11.186  <2e-16 ***
## as.factor(redu_ASA_C_TKA)else -0.7182     0.7949  -0.904    0.366
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 102.86  on 663  degrees of freedom
## AIC: 106.86
##
## Number of Fisher Scoring iterations: 7
```

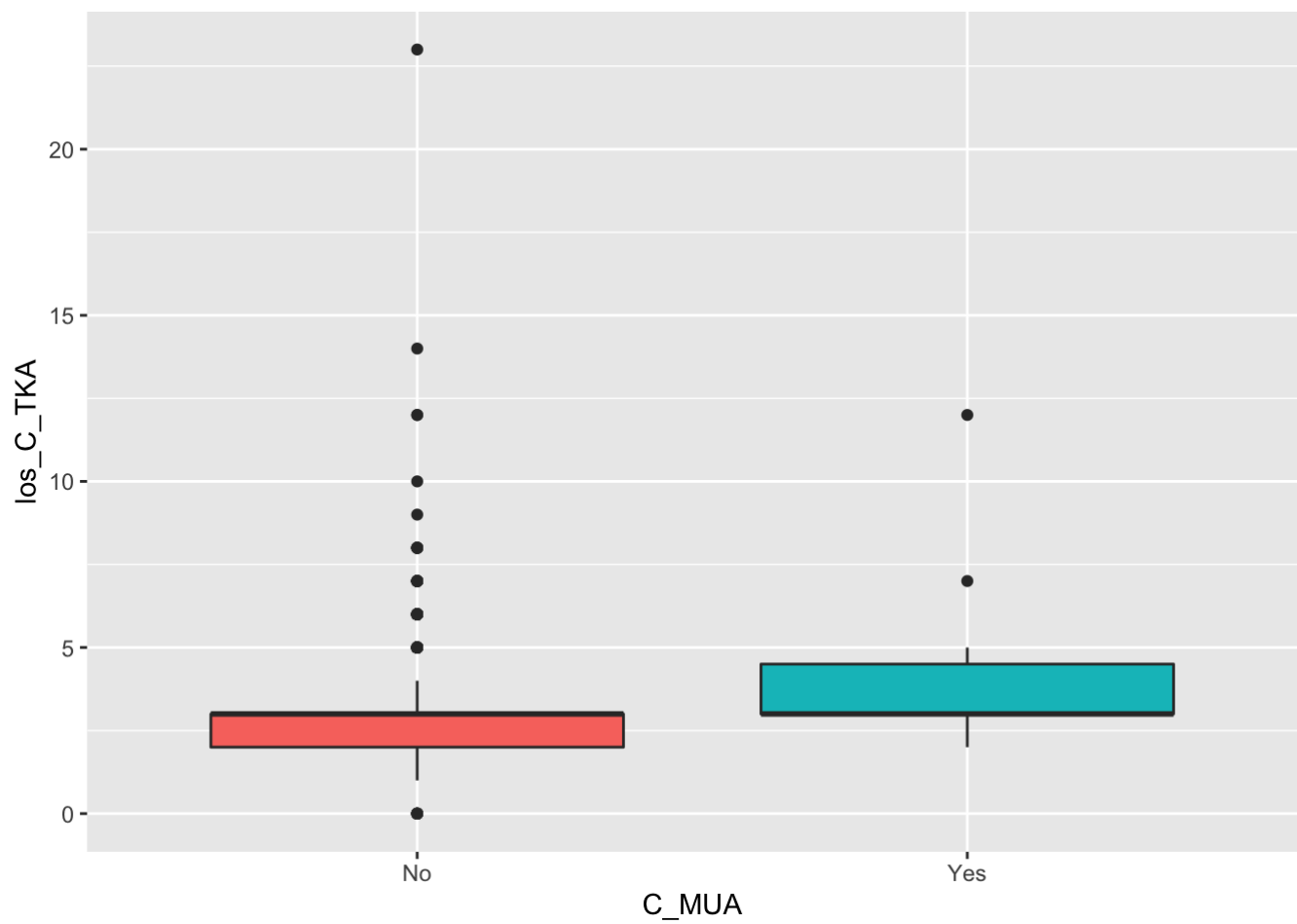
```
mua_glm1 = glm(C_MUA ~ MUA+ as.factor(redu_ASA_C_TKA), data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + as.factor(redu_ASA_C_TKA), family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5200  -0.1566  -0.1566  -0.1311   3.0858
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -4.3956     0.4425  -9.934 < 2e-16 ***
## MUAYes                2.4627     0.7521   3.274  0.00106 **
## as.factor(redu_ASA_C_TKA)else -0.3569     0.8289  -0.431  0.66672
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  95.053  on 662  degrees of freedom
## AIC: 101.05
##
## Number of Fisher Scoring iterations: 7
```

## (9) los

### a. EDA

Variable	N	No, N = 655 <sup>1</sup>	Yes, N = 10 <sup>1</sup>
<b>los_C_TKA</b>	665	3.00 (2.00, 3.00)	3.00 (3.00, 4.50)
<sup>1</sup> Median (IQR)			



b. regression model

```
mua_glm1 = glm(C_MUA ~ log_C_TKA, data= MUA_data1, family = "binomial")  
summary(mua_glm1)
```



```
##
## Call:
## glm(formula = C_MUA ~ los_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8965  -0.1665  -0.1665  -0.1524   2.9873
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.80709    0.48351  -9.942  <2e-16 ***
## los_C_TKA    0.17839    0.08877   2.010   0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 101.11  on 663  degrees of freedom
## AIC: 105.11
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA + los_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + los_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7002  -0.1429  -0.1429  -0.1316   3.0834
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.07478    0.53716  -9.447  < 2e-16 ***
## MUAYes       2.46618    0.73362   3.362 0.000775 ***
## los_C_TKA    0.16495    0.09424   1.750 0.080064 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  93.167  on 662  degrees of freedom
## AIC: 99.167
##
## Number of Fisher Scoring iterations: 7
```

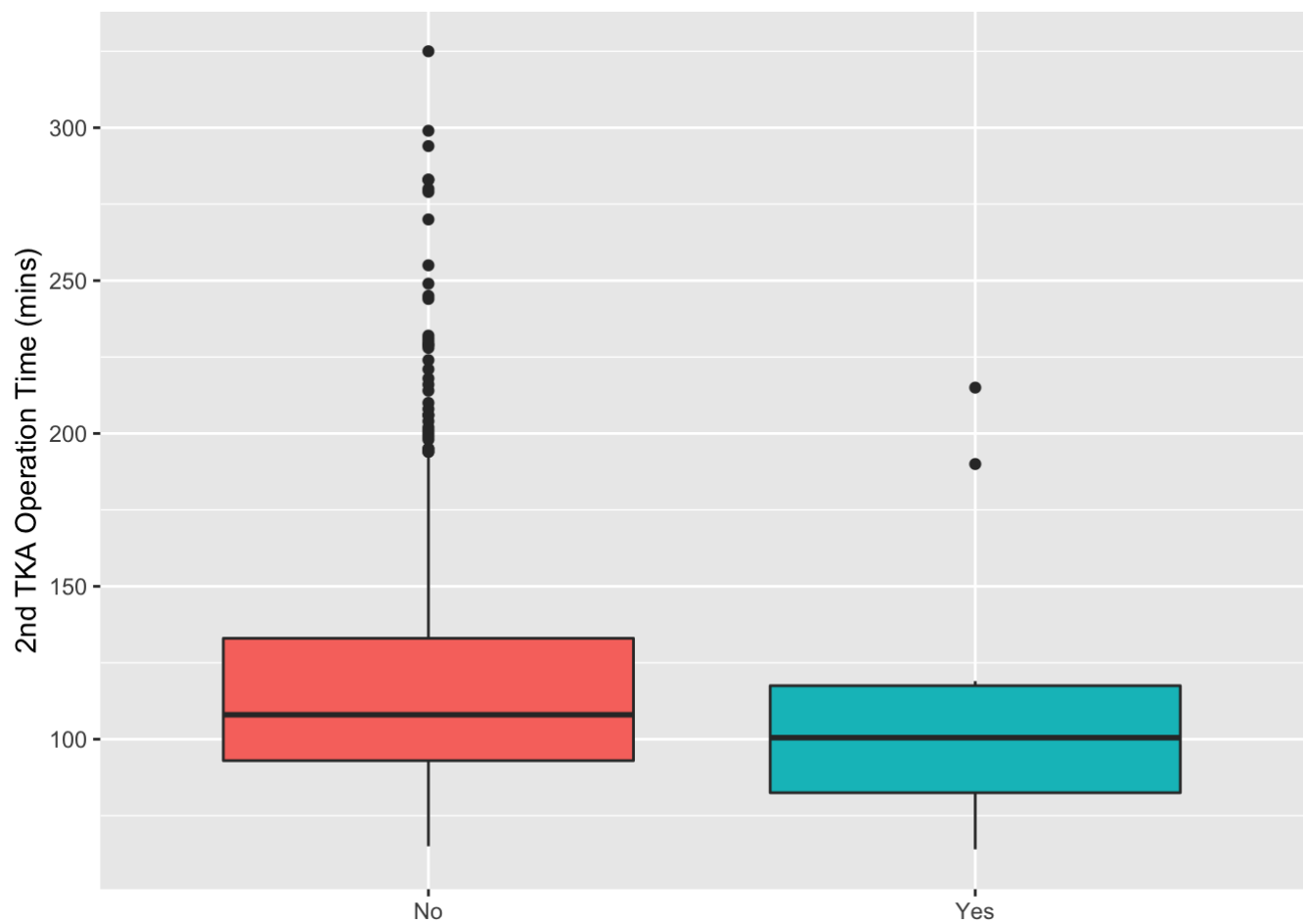
```
mua_glm1 = glm(C_MUA ~ (MUA + los_C_TKA)^2, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ (MUA + los_C_TKA)^2, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0343  -0.1506  -0.1483  -0.1483   3.0257
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.41180     0.84049  -5.249 1.53e-07 ***
## MUAYes         -0.61750     2.12504  -0.291   0.771
## los_C_TKA      -0.03105     0.24982  -0.124   0.901
## MUAYes:los_C_TKA  0.70005     0.43730   1.601   0.109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  88.715  on 661  degrees of freedom
## AIC: 96.715
##
## Number of Fisher Scoring iterations: 7
```

## (10) operation time

a. EDA

Variable	N	No, N = 655 <sup>†</sup>	Yes, N = 10 <sup>†</sup>
op_time_C_TKA	665	108 (93, 133)	100 (82, 118)
<sup>†</sup> Median (IQR)			



b. Regression model

```
mua_glm1 = glm(C_MUA ~ op_time_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ op_time_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.1893   -0.1809   -0.1762   -0.1688    3.0064
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.797912    1.088970  -3.488 0.000487 ***
## op_time_C_TKA -0.003304    0.009122  -0.362 0.717203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 103.65  on 663  degrees of freedom
## AIC: 107.65
##
## Number of Fisher Scoring iterations: 7
```

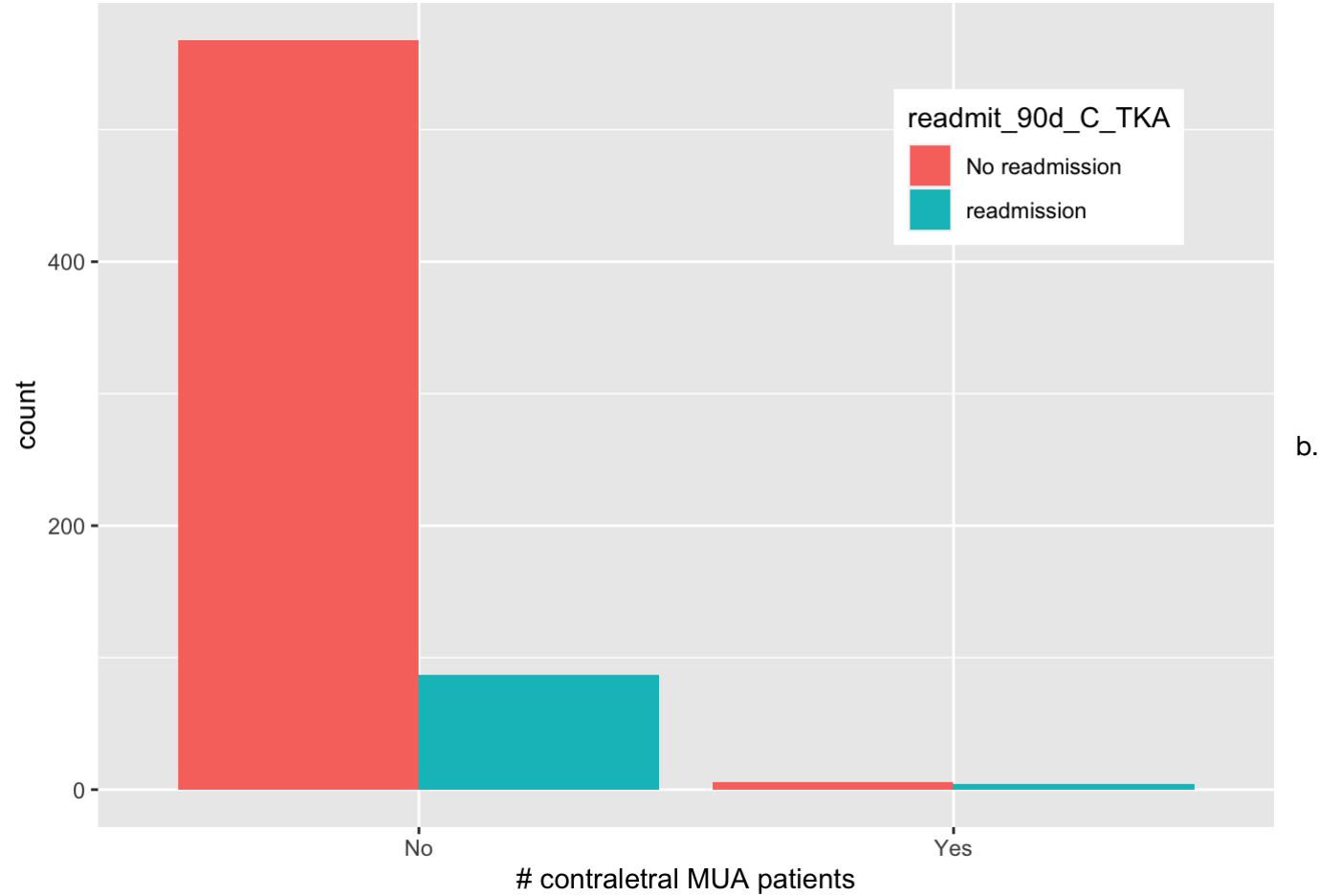
```
mua_glm1 = glm(C_MUA ~ MUA+op_time_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + op_time_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.5321   -0.1510   -0.1492   -0.1464    3.0063
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.337188    1.095228  -3.960 7.49e-05 ***
## MUAYes         2.547101    0.729099   3.493 0.000477 ***
## op_time_C_TKA -0.001434    0.008791  -0.163 0.870453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  95.221  on 662  degrees of freedom
## AIC: 101.22
##
## Number of Fisher Scoring iterations: 7
```

(11) readmission

a. EDA

Variable	N	No, N = 655 <sup>1</sup>	Yes, N = 10 <sup>1</sup>
readmit_90d_C_TKA	665		
0		568 (87%)	6 (60%)
1		87 (13%)	4 (40%)
<sup>1</sup> n (%)			

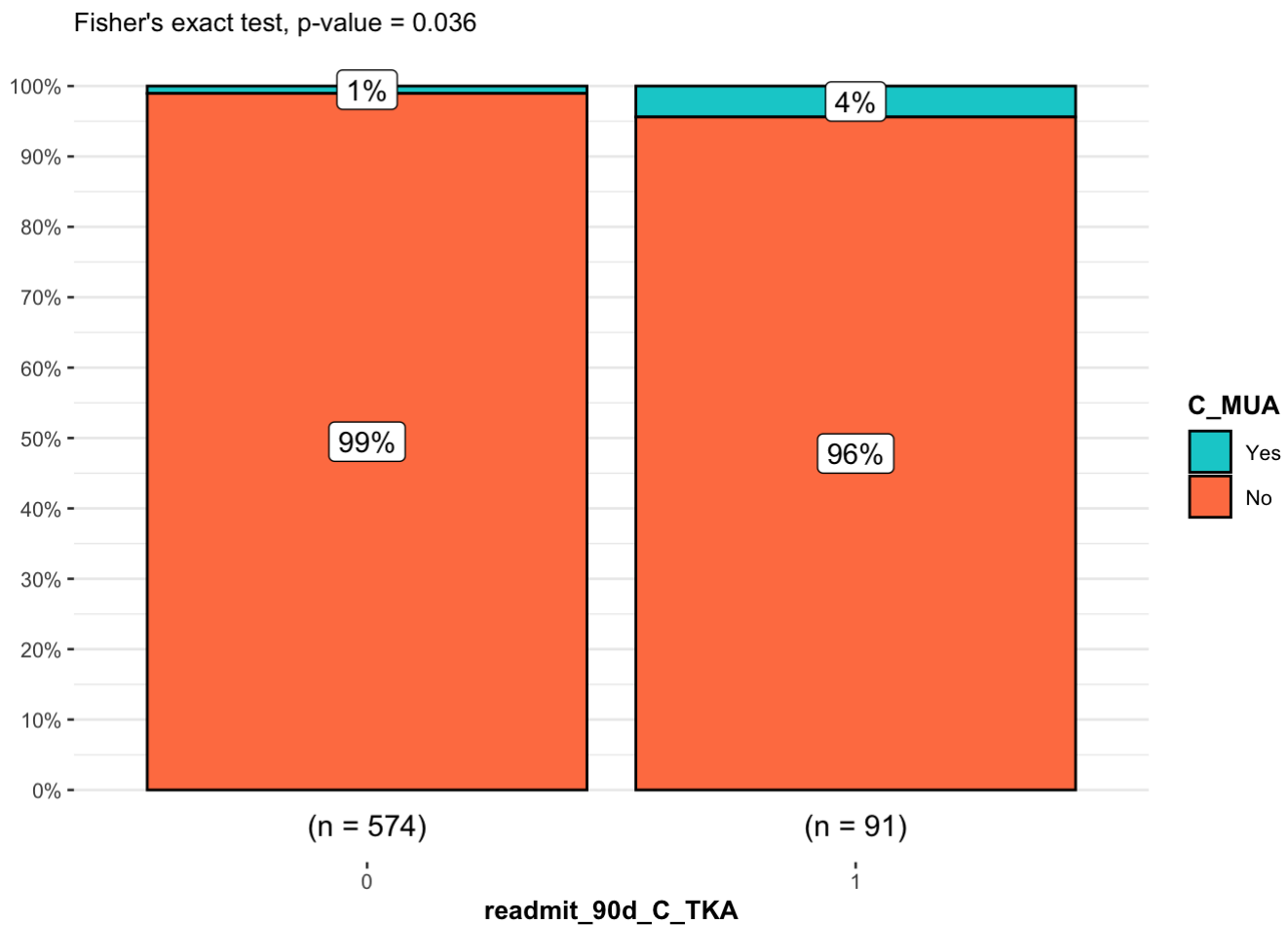


fisher

```
# fisher's exact test for count data
test<- fisher.test(table(MUA_data1$C_MUA, MUA_data1$readmit_90d_C_TKA))

# combine plot and statistical test with ggbarstats
ggbarstats(
  MUA_data1,C_MUA,readmit_90d_C_TKA,
  results.subtitle = FALSE,
  subtitle = paste0(
    "Fisher's exact test", ", ", p-value = ",
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))
  )
) + scale_fill_manual(values = c( "darkturquoise","coral"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```



### c. Regression model

```
mua_glm1 = glm(C_MUA ~ readmit_90d_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ readmit_90d_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2998  -0.1450  -0.1450  -0.1450   3.0202
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.5504     0.4104 -11.088  <2e-16 ***
## readmit_90d_C_TKA1  1.4707     0.6557   2.243  0.0249 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  99.486  on 663  degrees of freedom
## AIC: 103.49
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA+ readmit_90d_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

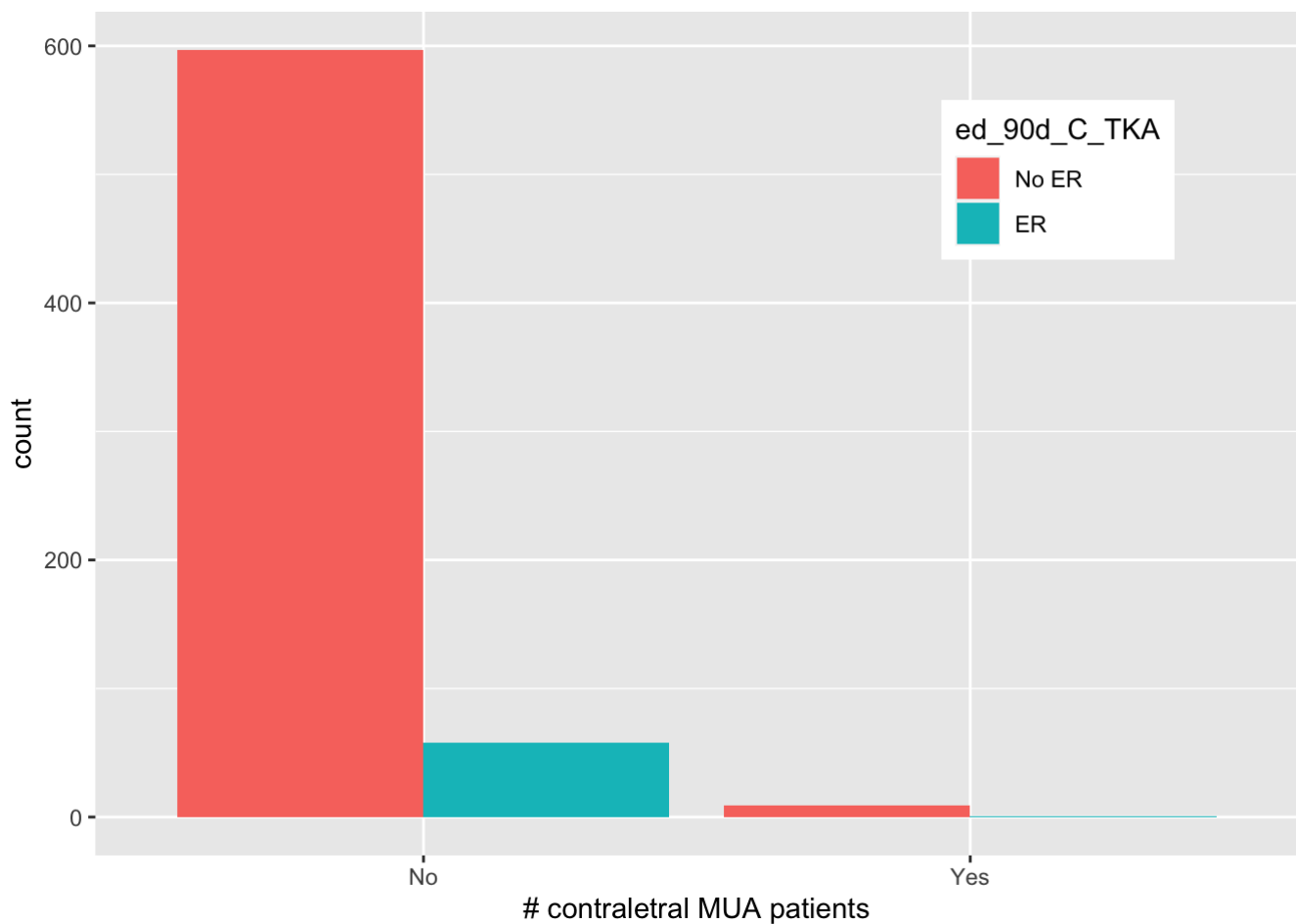
```
##
## Call:
## glm(formula = C_MUA ~ MUA + readmit_90d_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7985  -0.1257  -0.1257  -0.1257   3.1130
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.8375     0.4558 -10.614 < 2e-16 ***
## MUAYes         2.4747     0.7401   3.344 0.000826 ***
## readmit_90d_C_TKA1 1.3832     0.6737   2.053 0.040072 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance:  91.56  on 662  degrees of freedom
## AIC: 97.56
##
## Number of Fisher Scoring iterations: 7
```

### (13) ER visit within 90 days

#### a. EDA

Variable	N	No, N = 655 <sup>1</sup>	Yes, N = 10 <sup>1</sup>
ed_90d_C_TKA	665		
0		597 (91%)	9 (90%)
1		58 (8.9%)	1 (10%)
<sup>1</sup> n (%)			





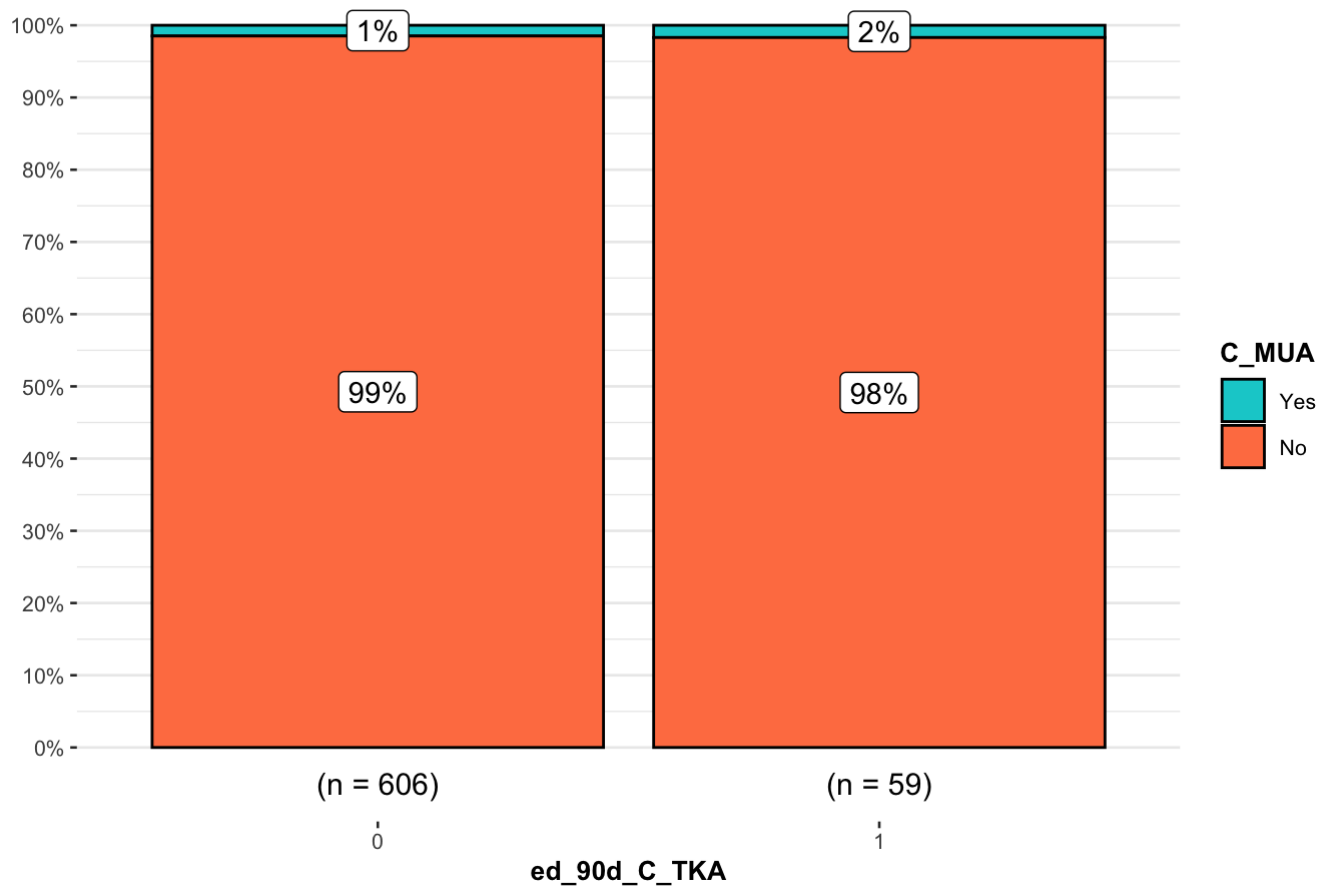
b. fisher

```
# fisher's exact test for count data
test<- fisher.test(table(MUA_data1$C_MUA, MUA_data1$ed_90d_C_TKA))

# combine plot and statistical test with ggbarstats
ggbarstats(
  MUA_data1,C_MUA,ed_90d_C_TKA,
  results.subtitle = FALSE,
  subtitle = paste0(
    "Fisher's exact test", ", p-value = ",
    ifelse(test$p.value < 0.001, "< 0.001", round(test$p.value, 3))
  )
) + scale_fill_manual(values = c( "darkturquoise","coral"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

Fisher's exact test, p-value = 0.608



### C. Regression model

```
mua_glm1 = glm(C_MUA ~ ed_90d_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ ed_90d_C_TKA, family = "binomial", data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1849  -0.1730  -0.1730  -0.1730   2.9016
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.1947     0.3358 -12.490  <2e-16 ***
## ed_90d_C_TKA1   0.1342     1.0630   0.126    0.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.79  on 664  degrees of freedom
## Residual deviance: 103.78  on 663  degrees of freedom
## AIC: 107.78
##
## Number of Fisher Scoring iterations: 7
```

```
mua_glm1 = glm(C_MUA ~ MUA+ ed_90d_C_TKA, data= MUA_data1, family = "binomial")
summary(mua_glm1)
```

```
##
## Call:
## glm(formula = C_MUA ~ MUA + ed_90d_C_TKA, family = "binomial",
##      data = MUA_data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5281  -0.1495  -0.1495  -0.1495   2.9997
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.4880     0.3873 -11.587  < 2e-16 ***
## MUAYes         2.5886     0.7355   3.520 0.000432 ***
## ed_90d_C_TKA1  -0.2401     1.1017  -0.218 0.827503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 103.793  on 664  degrees of freedom
## Residual deviance:  95.198  on 662  degrees of freedom
## AIC: 101.2
##
## Number of Fisher Scoring iterations: 7
```

(12) comorbidities

a. EDA

```
# summary
Comor_ratio_C_MUA <- apply(MUA_data1[MUA_data1$C_MUA=="Yes",c(82:89,91:101,103)],2,sum)/
10
Comor_ratio_NO_C_MUA<- round(apply(MUA_data1[MUA_data1$C_MUA!="Yes",c(82:89,91:101,10
3)],2,sum)/655,2)
cbind(Comor_ratio_C_MUA,Comor_ratio_NO_C_MUA)
```

##	Comor_ratio_C_MUA	Comor_ratio_NO_C_MUA
## blood_transfusion_C_TKA	0.5	0.05
## platelet_transfusion_C_TKA	0.0	0.00
## AIDS_C_TKA	0.0	0.01
## Malignancy_C_TKA	0.0	0.16
## Cerebrovascular_C_TKA	0.3	0.16
## COPD_C_TKA	0.1	0.23
## CHF_C_TKA	0.3	0.13
## Dementia_C_TKA	0.0	0.01
## Diabetes_no_cc_C_TKA	0.1	0.24
## Hemiplegia_C_TKA	0.1	0.01
## Metastatic_C_TKA	0.0	0.02
## Mild_Liver_C_TKA	0.1	0.17
## Moderate_Liver_C_TKA	0.1	0.02
## MI_C_TKA	0.0	0.10
## Peptic_Ulcer_C_TKA	0.0	0.05
## PVD_C_TKA	0.2	0.16
## CKD_C_TKA	0.1	0.11
## Rheumatic_C_TKA	0.0	0.07
## hematoma_C_TKA	0.0	0.00
## knee_infection_C_TKA	0.0	0.01

b. Lasso

```
#Lasso model - comorbidities
library(glmnet)
xlmatrix = data.matrix(MUA_data1[,c(82:89,91:101,103)])
colnames(xlmatrix)<- c(str_sub(colnames(xlmatrix),end=-7))

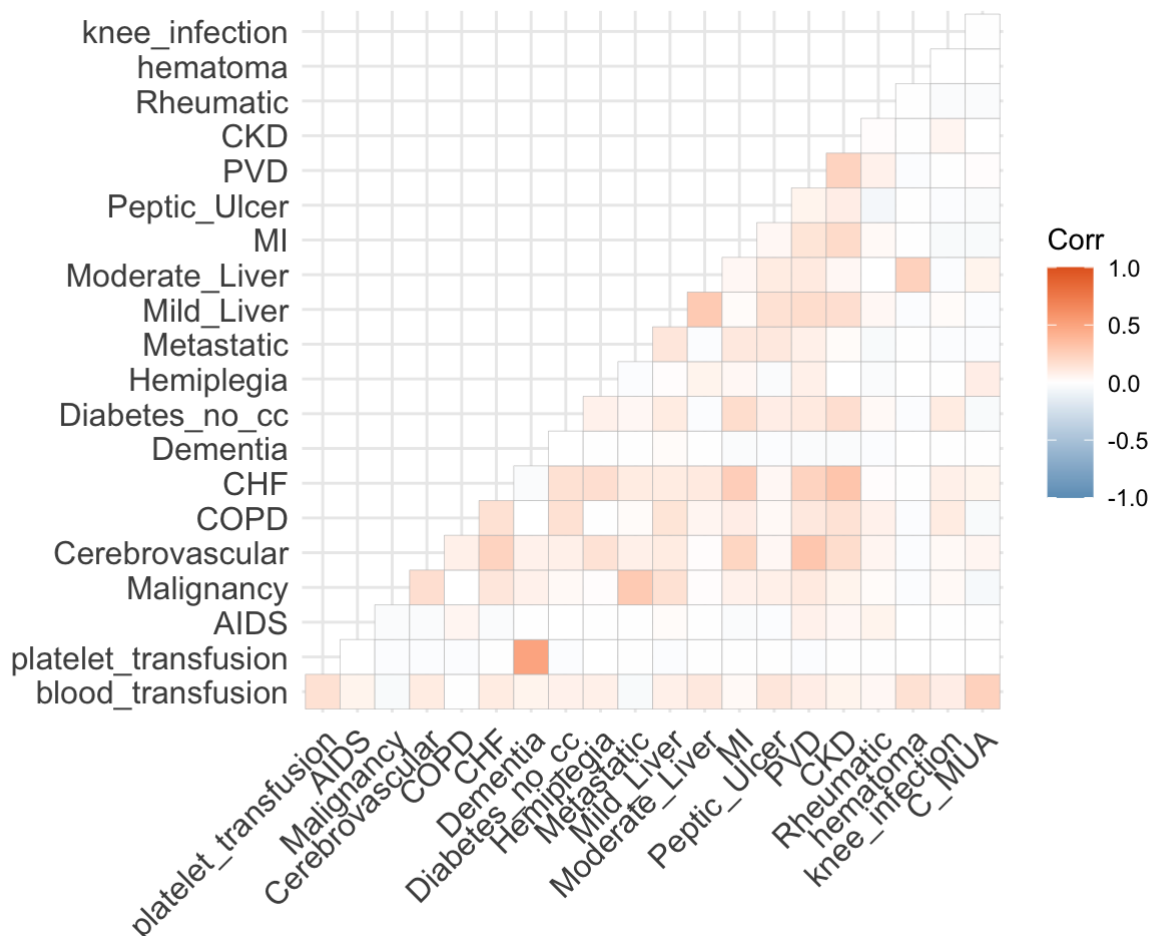
C_MUA = as.vector(MUA_data1$C_MUA_bi)

#Finding lambda
k = 5
set.seed(987)
cv.lasso<- cv.glmnet(xlmatrix, C_MUA, alpha=1, family = "binomial")
lasso.mod = glmnet(xlmatrix, C_MUA, alpha=1,family = "binomial",lambda=cv.lasso$lambda.min)
lasso.coef <- predict(lasso.mod, type = "coefficients", s=cv.lasso$lambda.min)
paste0(round(lasso.coef@x,4),"X",lasso.coef@i, collapse=" + ")
```

```
## [1] "-4.6106X0 + 2.3932X1 + 0.3195X10"
```

Lasso regression model picked blood transfusion and Hemiplegia

- correlation between diagnosed disease and contralateral MUA



The predictor that lasso picked matches with correlation between C\_MUA and diseases

So final model is..?

## (1) LASSO

a. model

```
library(glmnet)
names(MUA_data1)
```

```
## [1] "MUA_type"
## [2] "Group2"
## [3] "MUA"
## [4] "C_MUA"
## [5] "ID"
## [6] "sex"
## [7] "age"
## [8] "race"
## [9] "ethnicity"
## [10] "BMI"
## [11] "tobacco"
## [12] "death_date"
## [13] "Insurance"
## [14] "cpt code"
## [15] "surgery_date"
## [16] "Admission date"
## [17] "Discharge date"
## [18] "los"
## [19] "disch_disp"
## [20] "ASA"
## [21] "op_time"
## [22] "mua_count"
## [23] "blood_transfusion"
## [24] "platelet_transfusion"
## [25] "AIDS"
## [26] "Malignancy"
## [27] "Cerebrovascular"
## [28] "COPD"
## [29] "CHF"
## [30] "Dementia"
## [31] "Diabetes_cc"
## [32] "Diabetes_no_cc"
## [33] "Hemiplegia"
## [34] "Metastatic"
## [35] "Mild_Liver"
## [36] "Moderate_Liver"
## [37] "MI"
## [38] "Peptic_Ulcer"
## [39] "PVD"
## [40] "CKD"
## [41] "Rheumatic"
## [42] "hematoma"
## [43] "wound_infection"
## [44] "knee_infection"
## [45] "Readmission within 90 days (1=yes)"
## [46] "Readmission date"
## [47] "Reason for Readmission"
## [48] "ED visit within 90 days (1=yes)"
## [49] "ED visit date"
## [50] "Reason for ED Visit"
## [51] "Initial TKA Knee (right or left)"
## [52] "Pre-Op Total Knee arthroplasty range of motion"
```

```
## [53] "Date of most recent Preop Visit"
## [54] "varus/valgus preoperative (normal=0, varus=1, valgus=2)"
## [55] "Postoperative ROM (list flexion arc) 1"
## [56] "Date of first follow up appt"
## [57] "Postoperative ROM (list flexion arc) 2"
## [58] "Date of second follow up"
## [59] "Date of MUA"
## [60] "IF SAME KNEE HAS A 2ND MUA: Post MUA ROM ARC"
## [61] "DATE of follow up"
## [62] "DATE of 2nd MUA"
## [63] "Date of contralateral TKA"
## [64] "Contralateral side (left or right knee)"
## [65] "Pre-Op contralateral TKA ROM"
## [66] "Date of Most Recent Preop Visit_C_TKA"
## [67] "varus/valgus preoperative_C_TKA"
## [68] "Postoperative ROM (list flexion arc) 1_C_TKA"
## [69] "Date of f/u 1_C_TKA"
## [70] "Postoperative ROM (list flexion arc) 2_C_TKA"
## [71] "Date of f/u 2_C_TKA"
## [72] "Contralateral knee MUA?"
## [73] "Contralateral Knee MUA Date"
## [74] "age_C_TKA"
## [75] "bmi_C_TKA"
## [76] "tobacco_C_TKA"
## [77] "Insurance_C_TKA"
## [78] "los_C_TKA"
## [79] "disch_disp_C_TKA"
## [80] "ASA_C_TKA"
## [81] "op_time_C_TKA"
## [82] "blood_transfusion_C_TKA"
## [83] "platelet_transfusion_C_TKA"
## [84] "AIDS_C_TKA"
## [85] "Malignancy_C_TKA"
## [86] "Cerebrovascular_C_TKA"
## [87] "COPD_C_TKA"
## [88] "CHF_C_TKA"
## [89] "Dementia_C_TKA"
## [90] "Diabetes_cc_C_TKA"
## [91] "Diabetes_no_cc_C_TKA"
## [92] "Hemiplegia_C_TKA"
## [93] "Metastatic_C_TKA"
## [94] "Mild_Liver_C_TKA"
## [95] "Moderate_Liver_C_TKA"
## [96] "MI_C_TKA"
## [97] "Peptic_Ulcer_C_TKA"
## [98] "PVD_C_TKA"
## [99] "CKD_C_TKA"
## [100] "Rheumatic_C_TKA"
## [101] "hematoma_C_TKA"
## [102] "wound_infection_C_TKA"
## [103] "knee_infection_C_TKA"
## [104] "readmit_90d_C_TKA"
```



```
## [105] "readmit_date_C_TKA"
## [106] "ed_90d_C_TKA"
## [107] "ed_date_C_TKA"
## [108] "MUA_bi"
## [109] "C_MUA_bi"
## [110] "MUA_count_T"
## [111] "age_diff"
## [112] "los_total"
## [113] "op_time_total"
## [114] "date_diff"
## [115] "BTKA"
## [116] "Fac_age_C_TKA"
## [117] "Fac_age_TKA"
## [118] "redu_race"
## [119] "redu_tobacco"
## [120] "redu_tobacco_C_TKA"
## [121] "redu_ASA_C_TKA"
## [122] "redu_ASA"
## [123] "yeardiff"
## [124] "Fac_age_C_TKA2"
```

```
MUA_data2 = MUA_data1[,c(109,108,6,9,75:78, 80:89,91:104,106,116,118)]
MUA_data2$ASA_C_TKA<-as.factor(MUA_data2$ASA_C_TKA)
y = MUA_data2$C_MUA_bi

# Fac_age_C_TKA(116) is included instead of age_C_TKA(74), Diabetes_cc_C_TKA(90) is removed b/c it's duplicated with Diabetes_no_cc_C_TKA
# 108: MUA_bi, 6 : sex, 9: ethnicity, 75: bmi_C_TKA, 76: tobacco_C_TKA, 77: Insurance_C_TKA, 78: los_C_TKA, 80:ASA_C_TKA, 81: op_time_C_TKA, 82~103: comorbidities, 104:readmit_90d_C_TKA, 106:ed_90d_C_TKA,116:Fac_age_C_TKA, 118: redu_race

d <- as.data.frame(MUA_data2)
options(na.action="na.pass")
m <- model.matrix(C_MUA_bi ~ ., data=d)[,-1]

set.seed(1234)
cv.out = cv.glmnet(m, y, alpha=1,family="binomial")
bestlam = cv.out$lambda.min

lasso.mod = glmnet(m, y, alpha= 1, lambda = bestlam,family="binomial")
coef(lasso.mod)
```

```

## 45 x 1 sparse Matrix of class "dgCMatrix"
##                                     s0
## (Intercept)                      -4.6203440
## MUA_bi                           1.2846775
## sexM                             .
## ethnicityNon-Hispanic Origin    .
## bmi_C_TKA                        .
## tobacco_C_TKAPassive             .
## tobacco_C_TKAQuit                .
## tobacco_C_TKAYes                 .
## Insurance_C_TKAMedicare          .
## Insurance_C_TKAPrivate           .
## Insurance_C_TKAUninsured         .
## Insurance_C_TKAWork_Comp         0.9442249
## los_C_TKA                        .
## ASA_C_TKA2                       .
## ASA_C_TKA3                       .
## ASA_C_TKA4                       .
## op_time_C_TKA                    .
## blood_transfusion_C_TKA          2.0398585
## platelet_transfusion_C_TKA       .
## AIDS_C_TKA                       .
## Malignancy_C_TKA                 .
## Cerebrovascular_C_TKA            .
## COPD_C_TKA                       .
## CHF_C_TKA                        .
## Dementia_C_TKA                   .
## Diabetes_no_cc_C_TKA             .
## Hemiplegia_C_TKA                 .
## Metastatic_C_TKA                 .
## Mild_Liver_C_TKA                 .
## Moderate_Liver_C_TKA             .
## MI_C_TKA                         .
## Peptic_Ulcer_C_TKA               .
## PVD_C_TKA                        .
## CKD_C_TKA                        .
## Rheumatic_C_TKA                  .
## hematoma_C_TKA                   .
## wound_infection_C_TKA            .
## knee_infection_C_TKA             .
## readmit_90d_C_TKA1               .
## ed_90d_C_TKA1                    .
## Fac_age_C_TKA60s                 .
## Fac_age_C_TKAless50              .
## Fac_age_C_TKAover70s             .
## redu_raceOther                    .
## redu_raceWhite                    .

```

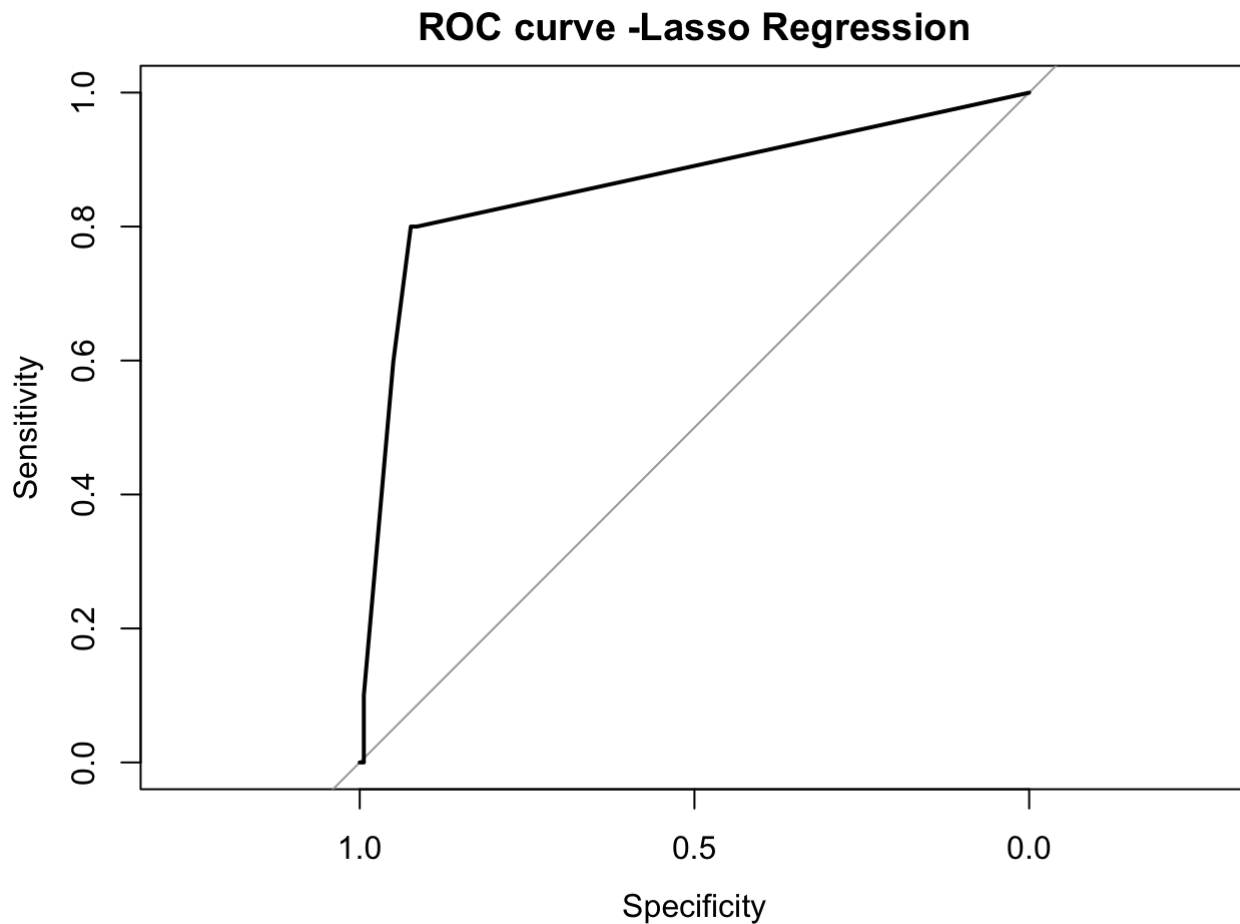
```
paste0(round(coef(lasso.mod)[x,4]),"X",coef(lasso.mod)[i, collapse=" + ")
```

```
## [1] "-4.6203X0 + 1.2847X1 + 0.9442X11 + 2.0399X17"
```

```
##### ROC
#lasso.mod = lasso mod using bestlam from cv.glmnet
#m = original model matrix
predict_fit = predict(lasso.mod, m, type = "response")

library(pROC)
roc_score=roc(predictor=as.vector(predict_fit), response = y )

#ROC Plot
plot(roc_score ,main = "ROC curve -Lasso Regression")
```

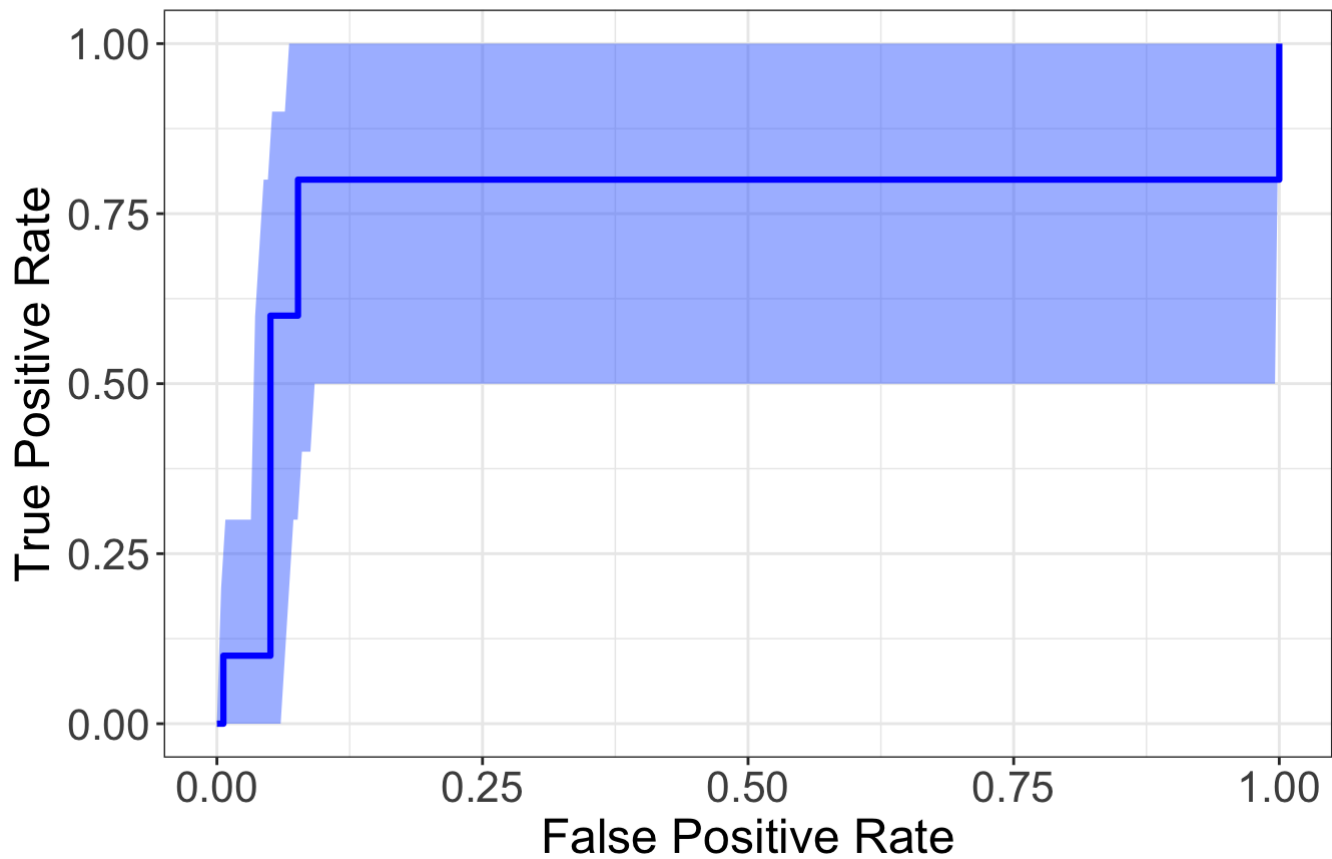


```
roc_score$auc #AUC score
```

```
## Area under the curve: 0.864
```

```
##### Bootstrapping AUC to find 95% CI
#Finding Bootstrapped AUC confidence intervals
library(fbroc)
boot_roc = boot.roc(as.numeric(predict_fit), as.logical(as.numeric(y)), n.boot = 10000)
plot(boot_roc)
```

# ROC Curve

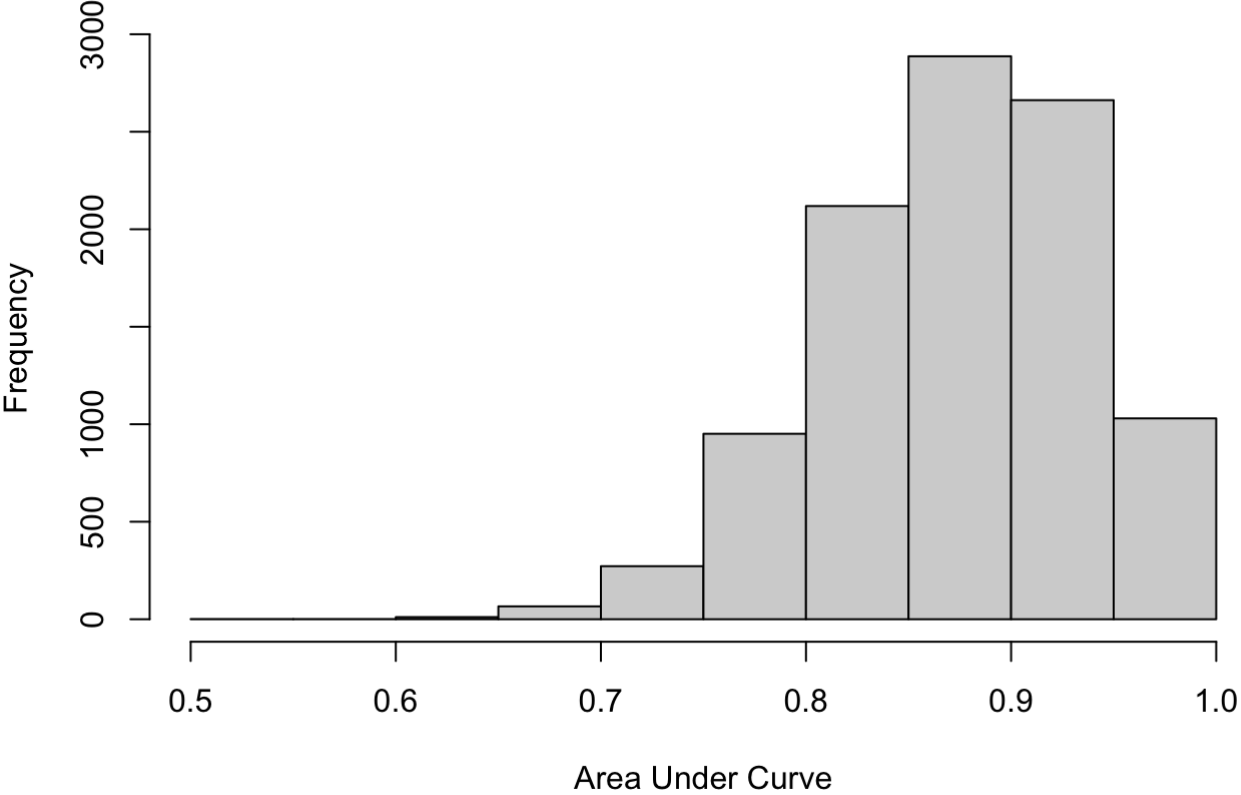


```
w=perf(boot_roc, "auc") #Measuring performance  
w #AUC confidence interval
```

```
##  
##  
##           Bootstrapped ROC performance metric  
##  
## Metric: AUC  
## Bootstrap replicates: 10000  
## Observed: 0.864  
## Std. Error: 0.065  
## 95% confidence interval:  
## 0.716 0.972
```

```
hist(w$boot.results, main = "Bootstrapped AUC Values", xlab = "Area Under Curve") #Histogram of bootstrapped AUC
```

# Bootstrapped AUC Values



$-4.6632 + 1.3594\text{MUA\_bi} + 1.2333\text{Insurance\_C\_TKAWork\_Comp} + 2.1177\text{blood\_transfusion\_C\_TKA}$

b. leave one out cross validation

```

library(glmnet)
#y = MUA_data1$C_MUA_bi
#MUA_data2 = MUA_data1[,c(108, 6, 9,75:78, 80:89,91:104,106,116,118)] # Fac_age_C_TKA(116) is included instead of age_C_TKA(74), Diabetes_cc_C_TKA(90) is removed b/c it's duplicated with Diabetes_no_cc_C_TKA
# 108: MUA_bi, 6 : sex, 9: ethnicity, 75: bmi_C_TKA, 76: tobacco_C_TKA, 77: Insurance_C_TKA, 78: los_C_TKA, 80:ASA_C_TKA, 81: op_time_C_TKA, 82~103: comorbidities, 104:readmit_90d_C_TKA, 106:ed_90d_C_TKA,116:Fac_age_C_TKA, 118: redu_race
#MUA_data2$ASA_C_TKA<-as.factor(MUA_data2$ASA_C_TKA)

#d <- data.frame(x=MUA_data2, y=y)
#m <- model.matrix(y ~ ., data=d)[,-1]

pred_prob <- c()
for(i in 1:nrow(d)){
y_train = d[-i,]$C_MUA_bi
x_train = m[-i,]

y_test = d[i,]$C_MUA_bi
x_test = m[i,]

set.seed(1234)
cv.out = cv.glmnet(x_train, y_train, alpha =1,family="binomial", nfolds=5)
bestlam = cv.out$lambda.min

lasso.mod = glmnet(x_train, y_train, alpha= 1, lambda = bestlam,family="binomial")
pred_prob[i] <- predict(lasso.mod,s = bestlam, newx = x_test,type="response")
}

pred_C_MUA<-cbind(pred_prob,y)
median(pred_C_MUA[y==0,1])

```

```
## [1] 0.007457973
```

```
median(pred_C_MUA[y==1,1])
```

```
## [1] 0.04405194
```

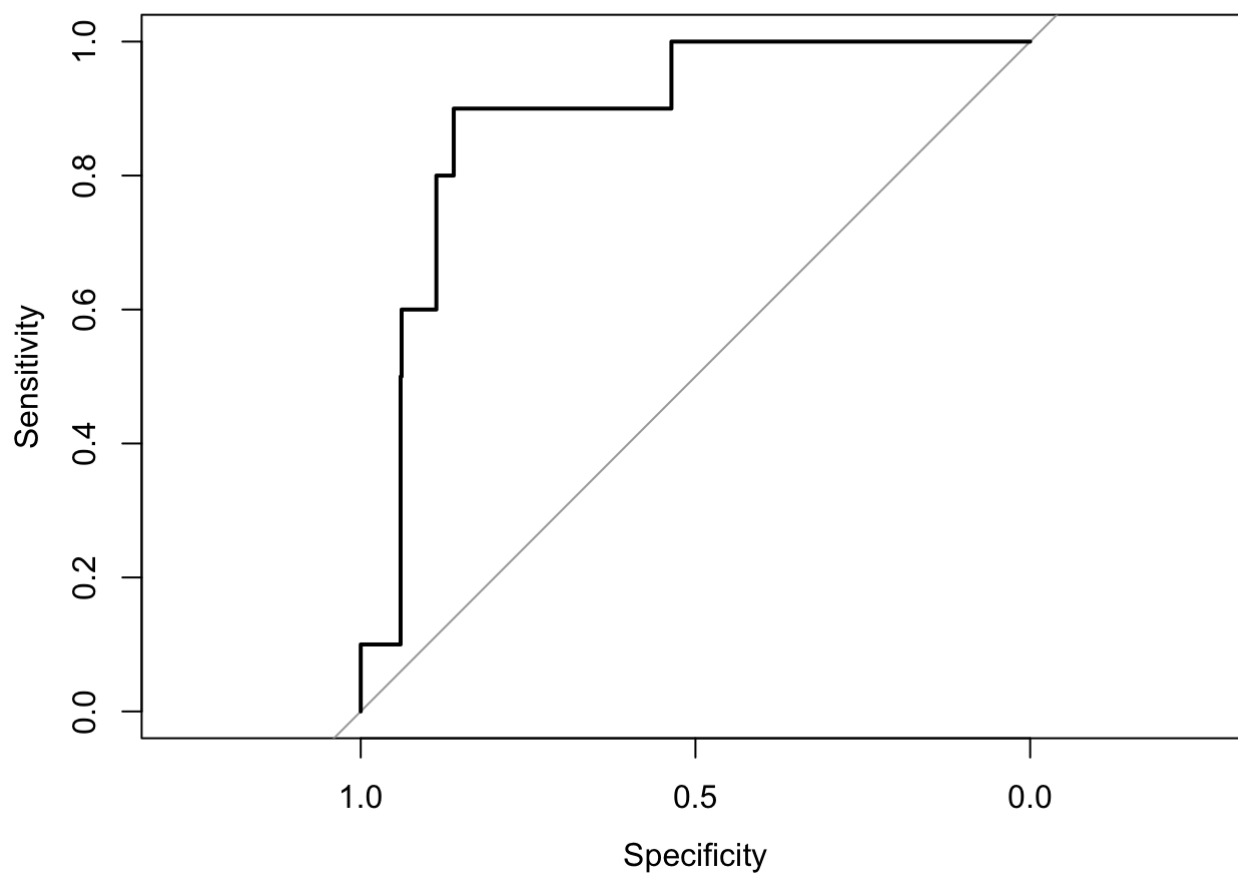
```

library(pROC)
roc_score_predictpower=roc(predictor=as.vector(pred_prob), response = y )

#ROC Plot
plot(roc_score_predictpower ,main ="ROC curve -Lasso Regression(LOOCV) ")

```

ROC curve -Lasso Regression(LOOCV)



```
roc_score_predictpower$auc #AUC score
```

```
## Area under the curve: 0.8872
```