

Lending Mortage Analysis

Exploratory Data Analysis

Marjorie Blanco, Joe Thompson, Haodi Tu

The data set contains 1989 records. The overall descriptive statistics:

```
##      Married      Meet credit history guidelines
##  No     : 678    0   : 171
##  Unknown:   3    1   :1816
##  Yes    :1308   666:    2
##
##
##
##  Other obligations as a percent of total income non-Hispanic Black
##  Min.   : 0.00                      No :1792
##  1st Qu.:28.00                     Yes: 197
##  Median :33.00
##  Mean   :32.39
##  3rd Qu.:37.00
##  Max.   :95.00
##  Hispanic      Male      Mortgage loan approved
##  No :1878    No     : 369    No : 244
##  Yes: 111   Unknown:  15    Yes:1745
##                  Yes     :1605
##
##
##
##  Loan amount/purchase price           Race
##  Min.   : 2.105          Hispanic       : 111
##  1st Qu.: 70.000         non-Hispanic Black: 197
##  Median : 80.000         non-Hispanic White:1681
##  Mean   : 77.064
##  3rd Qu.: 89.894
##  Max.   :257.143
```

Descriptive statistics by Race:

```
## $Hispanic
##      MARRIED    GDLIN      OBRAT      MALE      APPROVE
##  No     :31    0   :16    Min.   :14.60    No     :22    No :26
##  Unknown: 1    1   :95   1st Qu.:29.00  Unknown: 2    Yes:85
##  Yes    :79   666: 0   Median :33.00    Yes    :87
##                  Mean   :33.46
##                  3rd Qu.:38.45
##                  Max.   :62.00
##      LOANPRC            RACE
##  Min.   : 39.39  Hispanic       :111
##  1st Qu.: 80.00  non-Hispanic Black: 0
##  Median : 89.39  non-Hispanic White: 0
##  Mean   : 85.17
##  3rd Qu.: 90.42
##  Max.   :162.63
```

```

##  

## $`non-Hispanic Black`  

##      MARRIED      GDLIN       OBRAT        MALE      APPROVE  

##  No   : 76    0   : 53   Min.   : 5.60   No   : 51   No  : 64  

## Unknown:  0    1   :144   1st Qu.:31.00  Unknown:  2   Yes:133  

## Yes   :121   666:  0   Median  :35.00   Yes   :144  

##                               Mean   :34.94  

##                               3rd Qu.:38.90  

##                               Max.   :63.00  

##  

##      LOANPRC          RACE  

##  Min.   : 28.99  Hispanic   : 0  

##  1st Qu.: 80.00  non-Hispanic Black:197  

##  Median  : 87.02  non-Hispanic White: 0  

##  Mean   : 83.97  

##  3rd Qu.: 90.24  

##  Max.   :255.52  

##  

## $`non-Hispanic White`  

##      MARRIED      GDLIN       OBRAT        MALE      APPROVE  

##  No   : 571    0   : 102   Min.   : 0.00   No   : 296   No  : 154  

## Unknown:  2    1   :1577  1st Qu.:27.60  Unknown:  11   Yes:1527  

## Yes   :1108   666:  2   Median  :32.50   Yes   :1374  

##                               Mean   :32.02  

##                               3rd Qu.:36.50  

##                               Max.   :95.00  

##  

##      LOANPRC          RACE  

##  Min.   : 2.105  Hispanic   : 0  

##  1st Qu.: 68.182 non-Hispanic Black: 0  

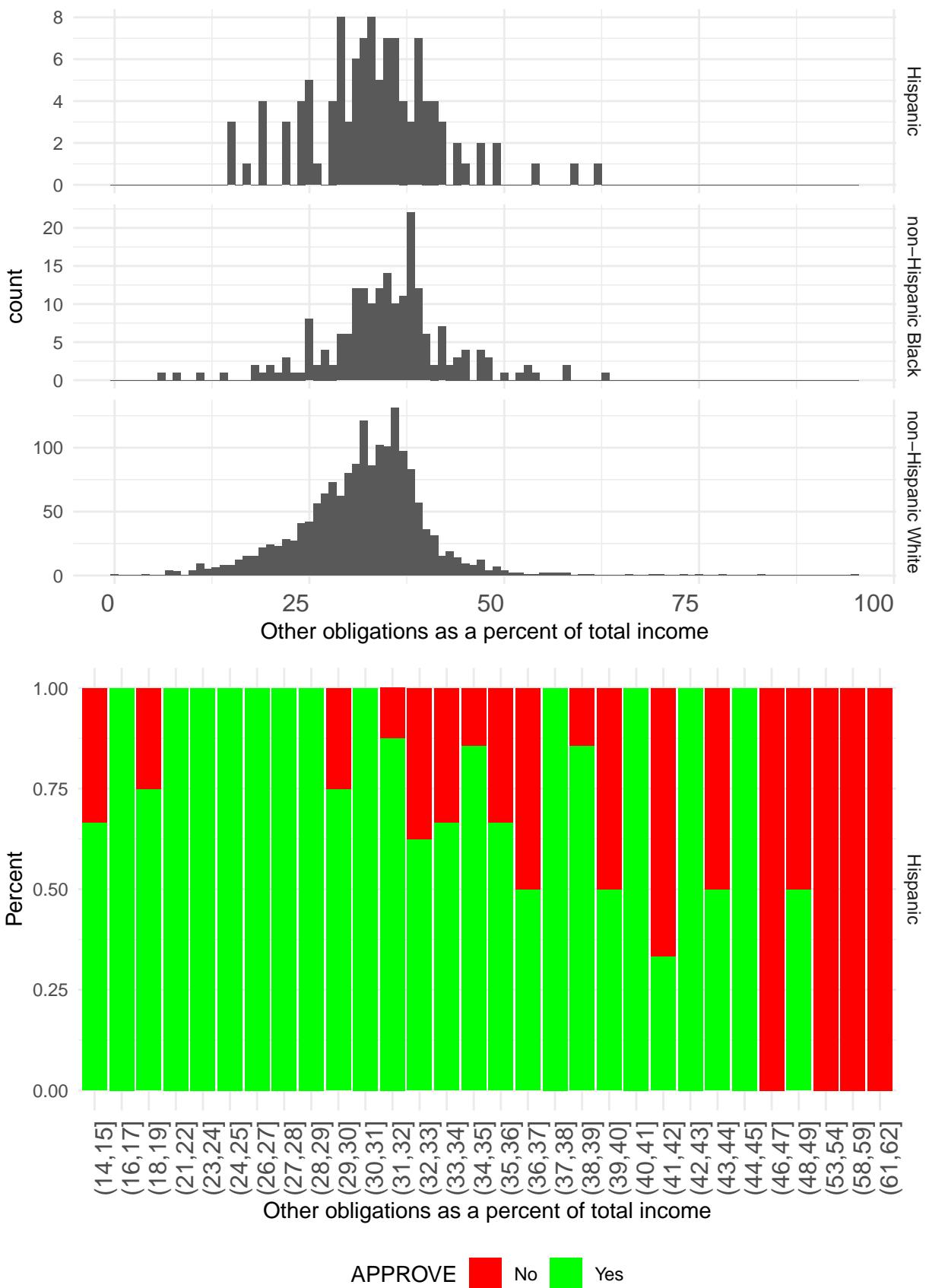
##  Median  : 79.888 non-Hispanic White:1681  

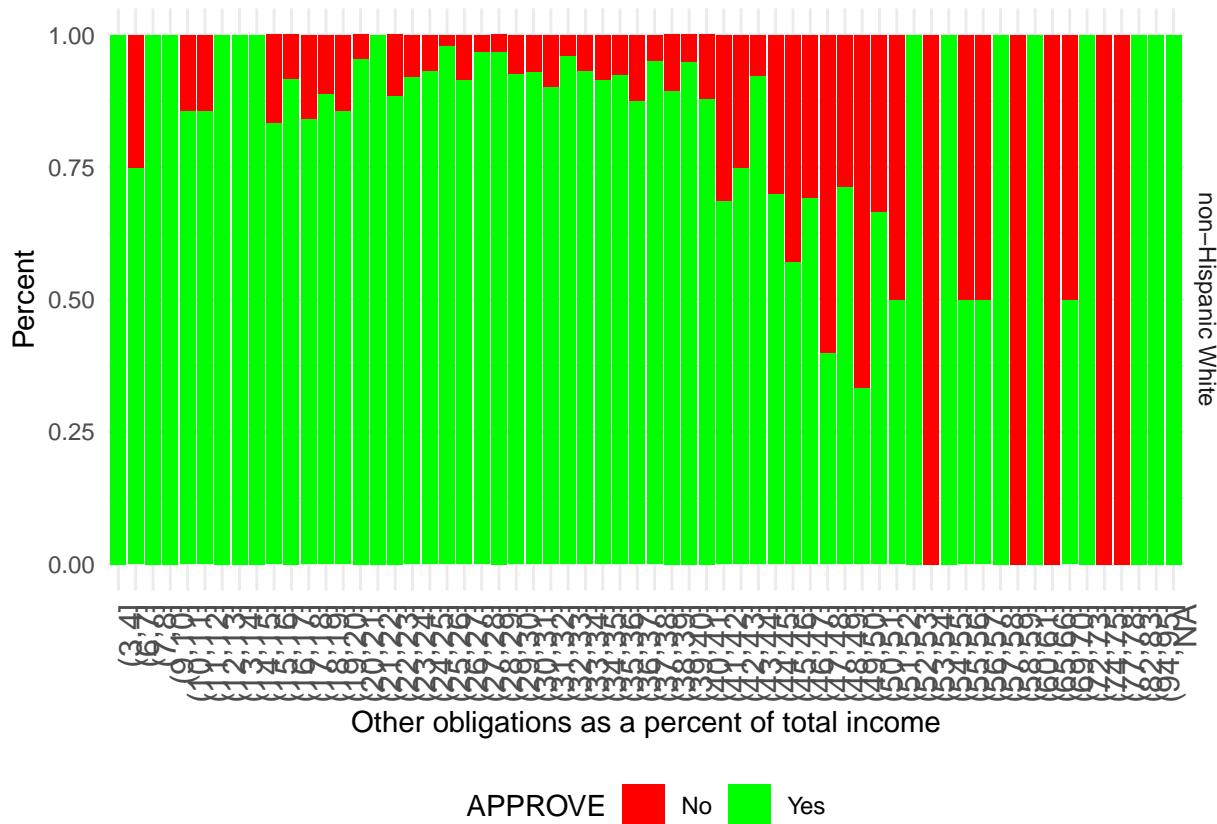
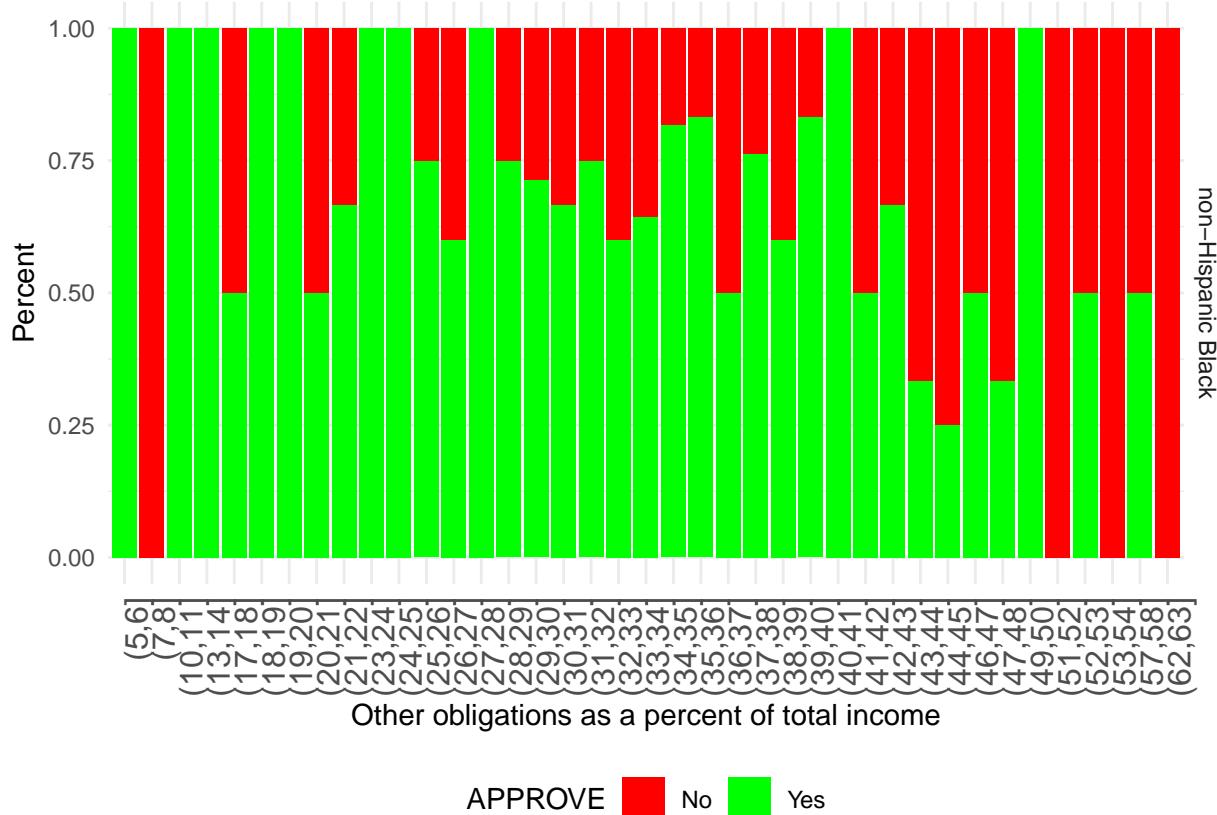
##  Mean   : 75.719  

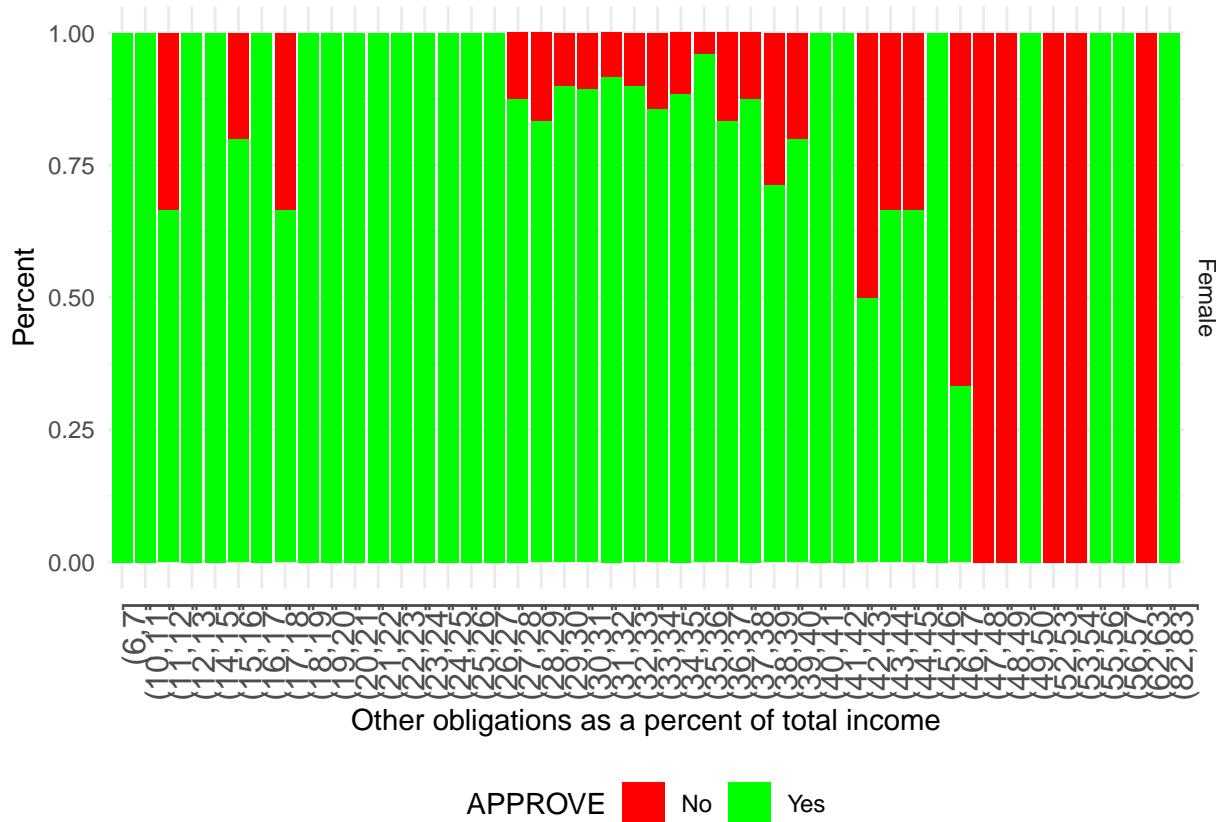
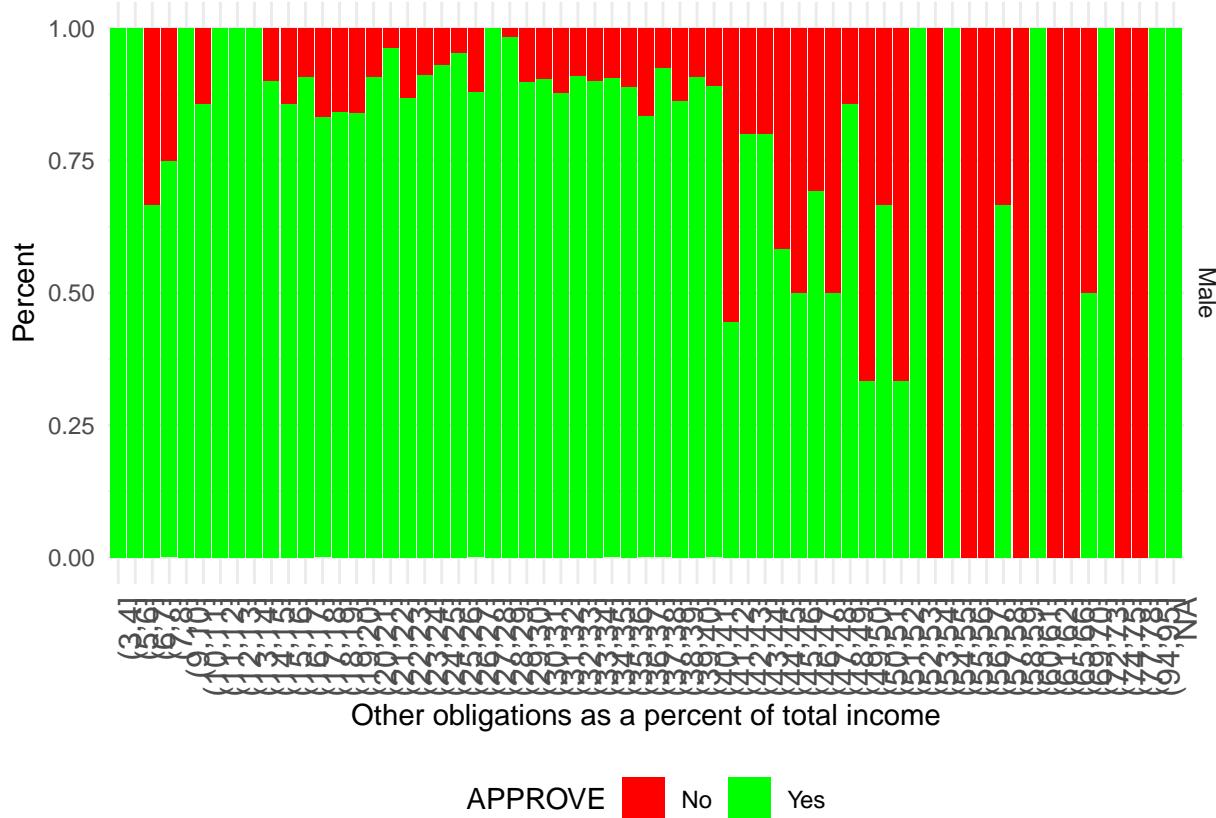
##  3rd Qu.: 89.623  

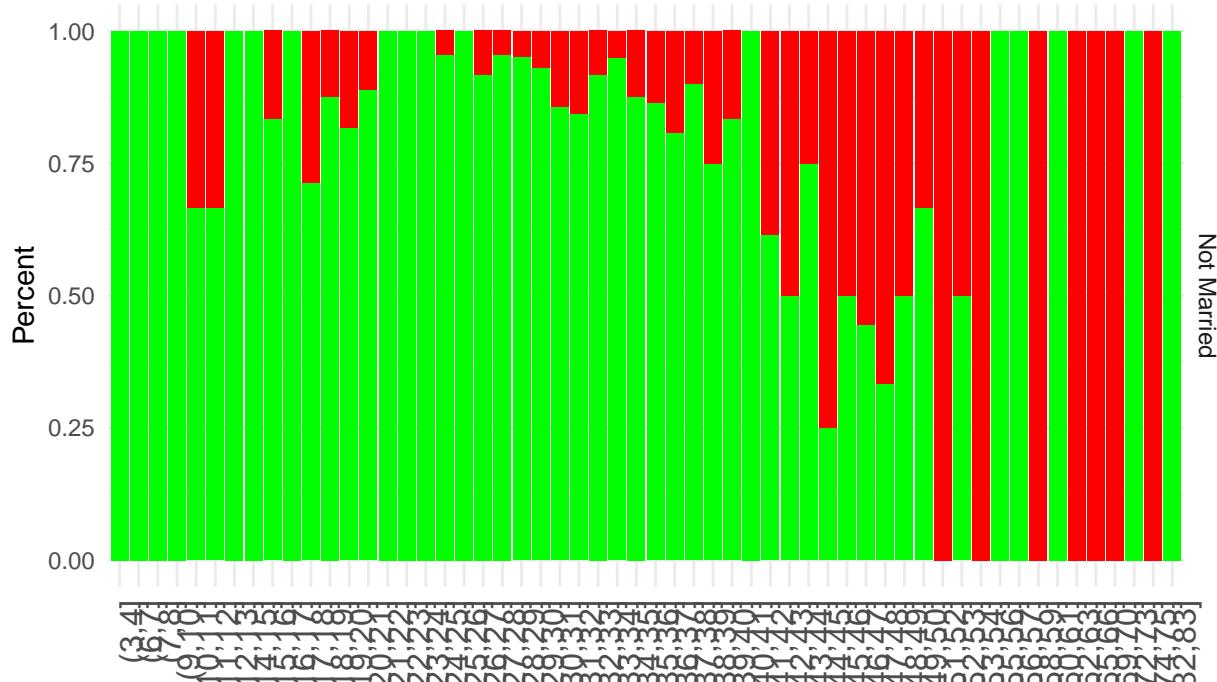
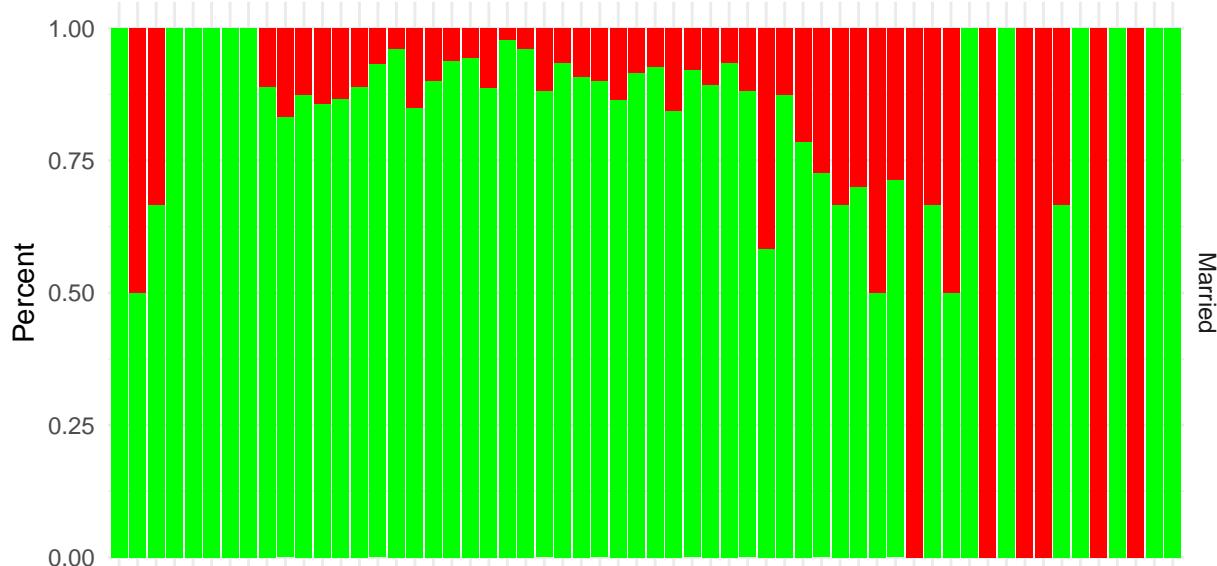
##  Max.   :257.143

```

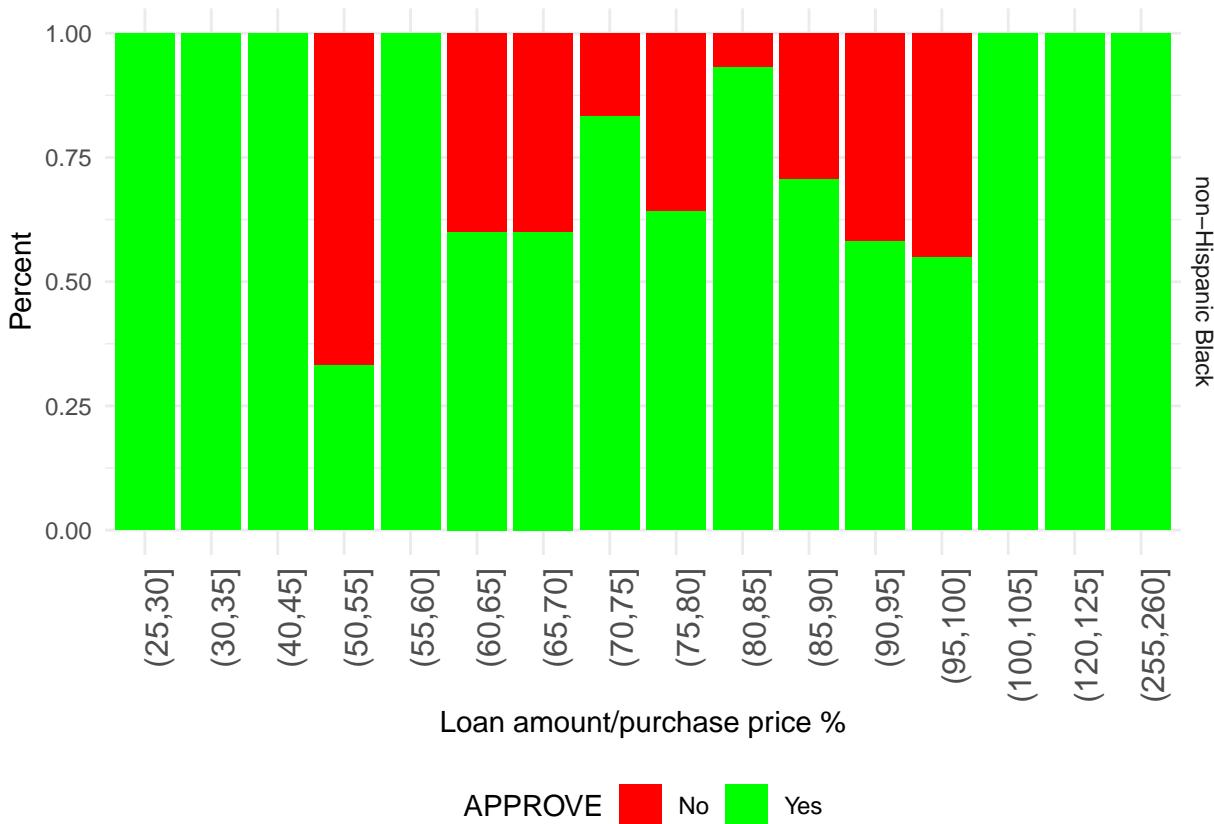
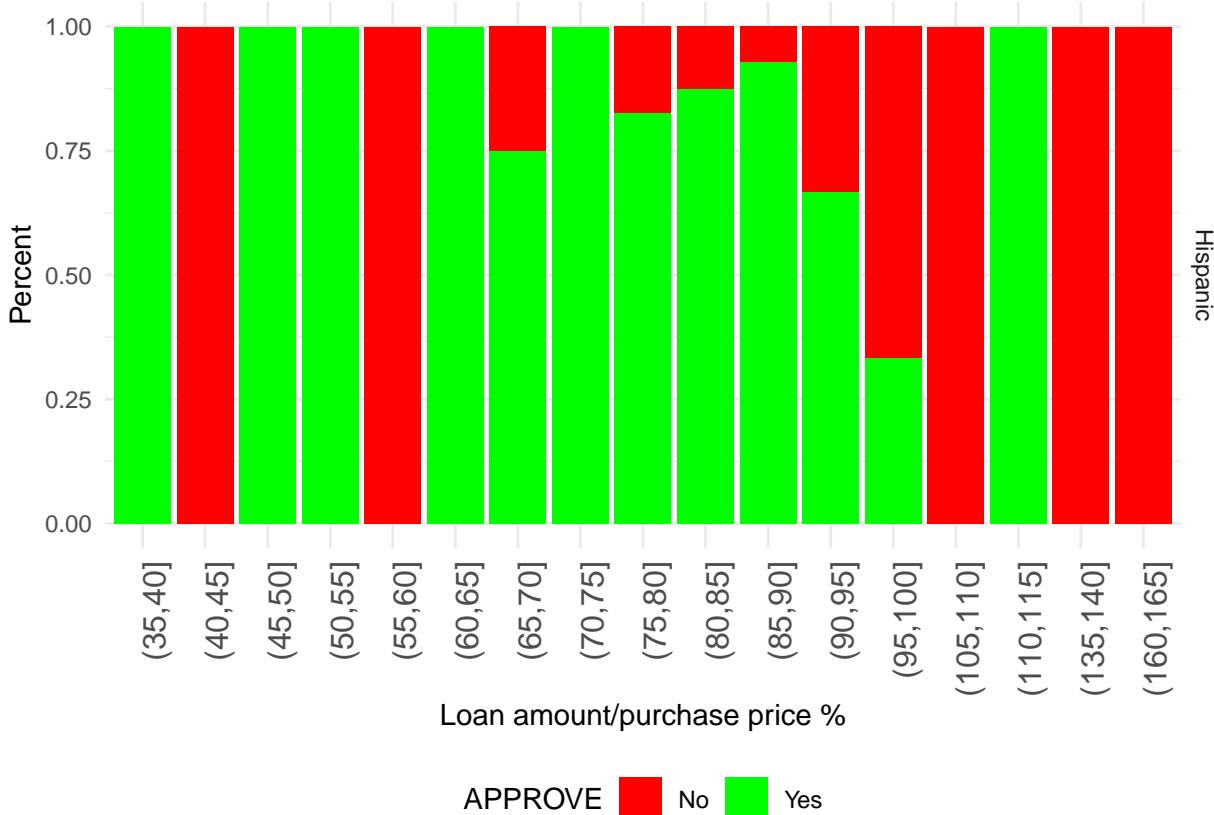


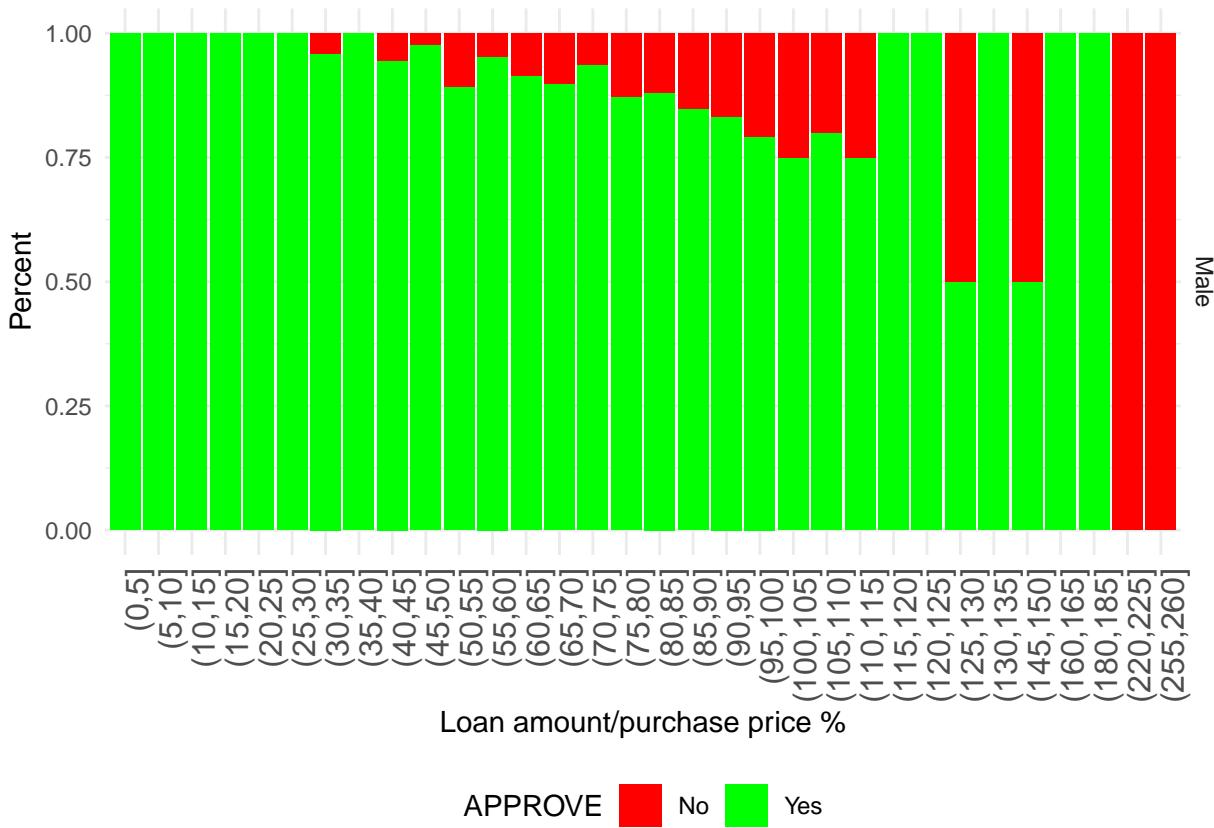
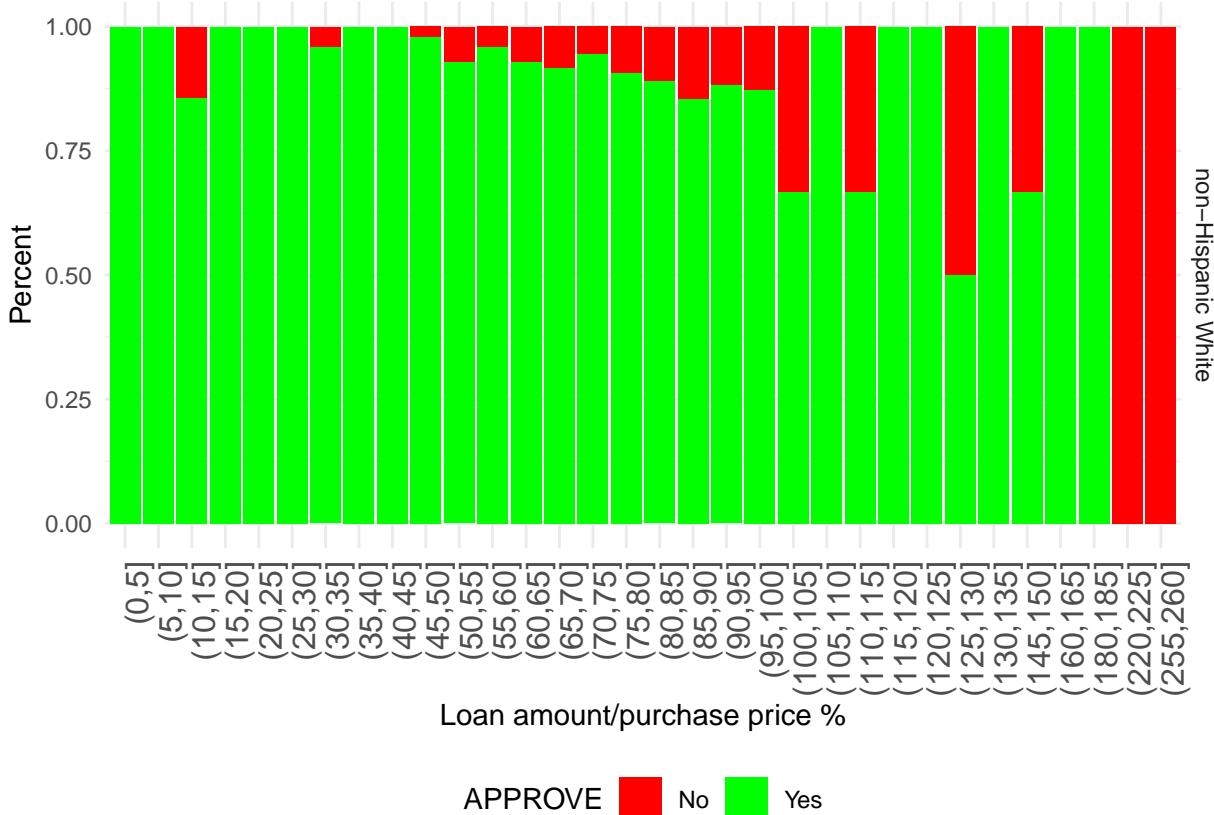


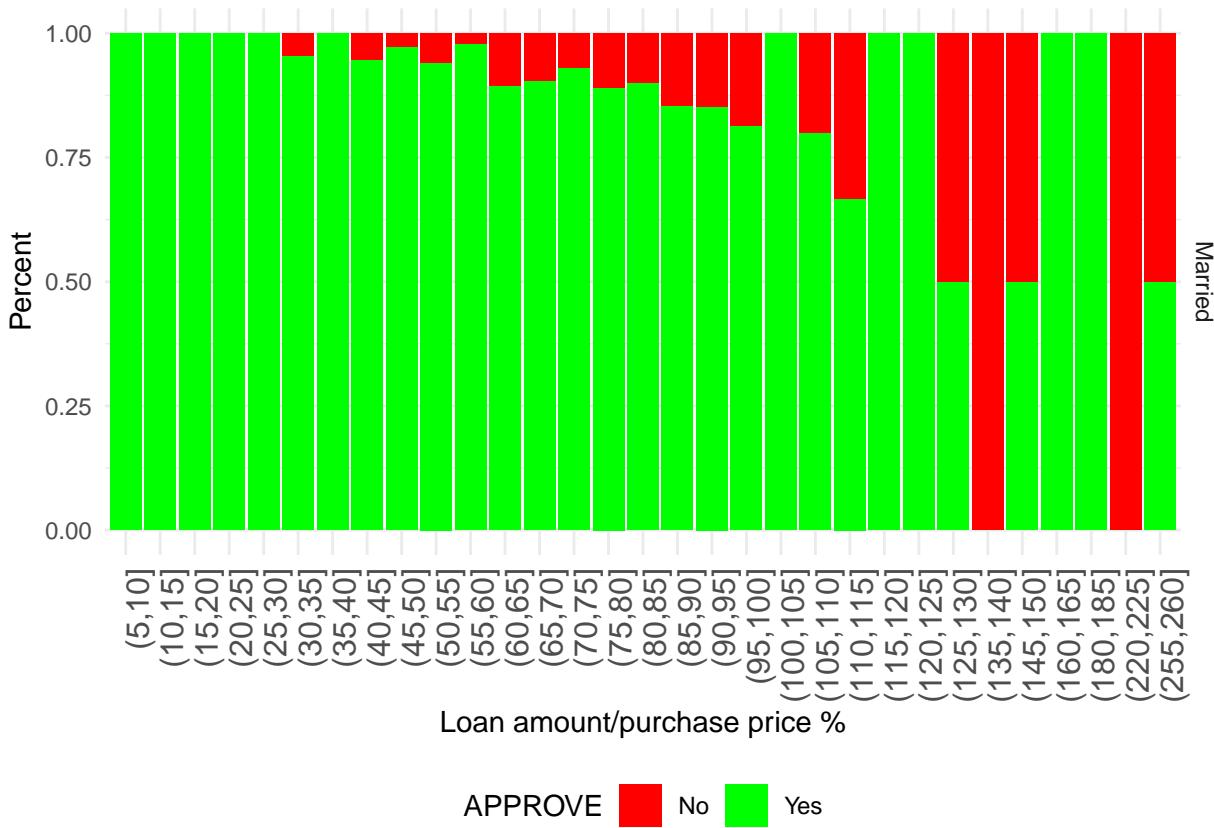
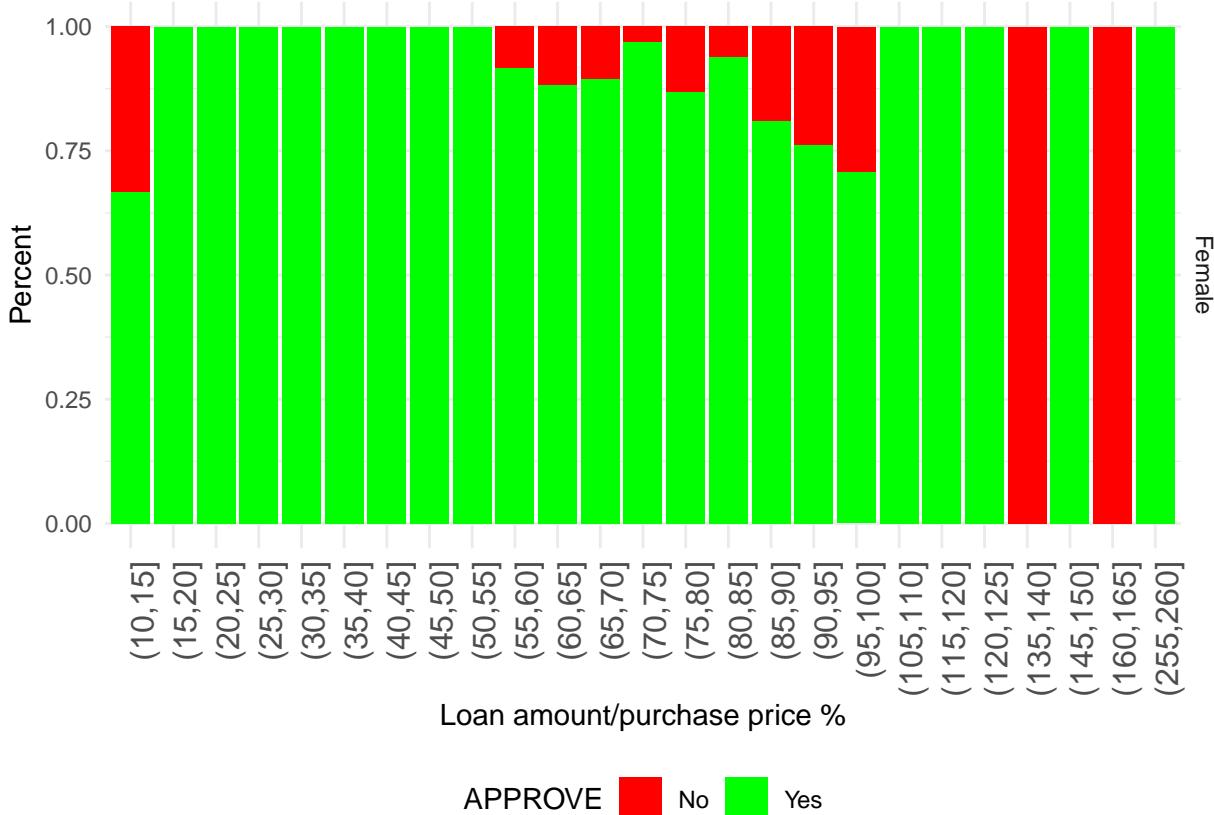




APPROVE No Yes









Descriptive statistics by Marital Status:

```

## $No
##      MARRIED    GDLIN        OBRAT        MALE      APPROVE
##  No     :678    0   : 64   Min.    : 4.00   No     :252   No  :102
##  Unknown:  0    1   :614   1st Qu.:28.00  Unknown:  7   Yes:576
##  Yes     : 0   666:  0   Median  :33.00   Yes    :419
##                               Mean   :32.74
##                               3rd Qu.:37.00
##                               Max.   :83.00
##      LOANPRC          RACE
##  Min.   : 2.105   Hispanic       : 31
##  1st Qu.: 72.426  non-Hispanic Black: 76
##  Median : 80.000  non-Hispanic White:571
##  Mean   : 77.967
##  3rd Qu.: 89.978
##  Max.   :162.626
##
## $Unknown
##      MARRIED    GDLIN        OBRAT        MALE      APPROVE    LOANPRC
##  No     :0    0   :0   Min.    :13.0   No     :1   No  :0   Min.   : 86.96
##  Unknown:3   1   :3   1st Qu.:23.3  Unknown:0   Yes:3   1st Qu.: 88.62
##  Yes     :0   666:0   Median  :33.6   Yes    :2
##                               Mean   :27.2
##                               3rd Qu.:34.3
##                               Max.   :35.0
##      RACE
##  Hispanic       :1
##  non-Hispanic Black:0

```

```

## non-Hispanic White:2
##
##
##
## $Yes
##      MARRIED      GDLIN       OBRAT        MALE      APPROVE
##  No      : 0  0 : 107  Min.   : 0.00  No     : 116  No  : 142
## Unknown: 0  1 :1199  1st Qu.:28.00  Unknown:  8  Yes:1166
## Yes     :1308  666:  2  Median  :33.00  Yes    :1184
##                               Mean   :32.22
##                               3rd Qu.:37.00
##                               Max.   :95.00
##      LOANPRC          RACE
##  Min.   : 8.772  Hispanic      : 79
##  1st Qu.: 68.857 non-Hispanic Black: 121
##  Median : 80.000 non-Hispanic White:1108
##  Mean   : 76.547
##  3rd Qu.: 89.866
##  Max.   :257.143

```

Descriptive statistics by Gender:

```

## $No
##      MARRIED      GDLIN       OBRAT        MALE      APPROVE
##  No      :252  0 : 31  Min.   : 6.99  No     :369  No  : 50
## Unknown: 1  1 :338  1st Qu.:28.00  Unknown:  0  Yes:319
## Yes     :116  666:  0  Median  :33.00  Yes    : 0
##                               Mean   :32.64
##                               3rd Qu.:37.00
##                               Max.   :83.00
##      LOANPRC          RACE
##  Min.   : 11.01  Hispanic      : 22
##  1st Qu.: 70.83 non-Hispanic Black: 51
##  Median : 80.00 non-Hispanic White:296
##  Mean   : 77.66
##  3rd Qu.: 90.00
##  Max.   :255.52
##
## $Unknown
##      MARRIED      GDLIN       OBRAT        MALE      APPROVE
##  No      :7  0 : 0  Min.   :24.00  No     : 0  No  : 0
## Unknown:0  1 :15  1st Qu.:29.95  Unknown:15  Yes:15
## Yes     :8  666: 0  Median  :34.50  Yes    : 0
##                               Mean   :33.33
##                               3rd Qu.:37.65
##                               Max.   :40.30
##      LOANPRC          RACE
##  Min.   :39.39  Hispanic      : 2
##  1st Qu.:74.93 non-Hispanic Black: 2
##  Median :75.42 non-Hispanic White:11
##  Mean   :75.59
##  3rd Qu.:80.43
##  Max.   :92.90
##
## $Yes

```

```

##      MARRIED      GDLIN        OBRAT        MALE      APPROVE
##  No   : 419    0 : 140    Min.   : 0.00    No    : 0    No : 194
## Unknown:  2     1 :1463   1st Qu.:28.00  Unknown:  0    Yes:1411
## Yes   :1184    666:   2   Median :33.00   Yes   :1605
##                               Mean   :32.32
##                               3rd Qu.:37.00
##                               Max.   :95.00
##      LOANPRC          RACE
##  Min.   : 2.105  Hispanic       : 87
##  1st Qu.: 69.655 non-Hispanic Black: 144
##  Median : 80.000 non-Hispanic White:1374
##  Mean   : 76.942
##  3rd Qu.: 89.881
##  Max.   :257.143

```

There are 3 records are missing married (MARRIED) field.

ID	MARRIED	GDLIN	OBRAT	BLACK	HISPAN	MALE	APPROVE	LOANPRC	RACE
356	Unknown	1	35.0	No	Yes	Yes	Yes	86.95652	Hispanic
759	Unknown	1	33.6	No	No	Yes	Yes	90.28571	non-Hispanic White
1392	Unknown	1	13.0	No	No	No	Yes	117.24140	non-Hispanic White

There are 3 records are missing married (GDLIN) field.

ID	MARRIED	GDLIN	OBRAT	BLACK	HISPAN	MALE	APPROVE	LOANPRC	RACE
881	Yes	666	35	No	No	Yes	Yes	75.82939	non-Hispanic White
1229	Yes	666	26	No	No	Yes	Yes	100.00000	non-Hispanic White

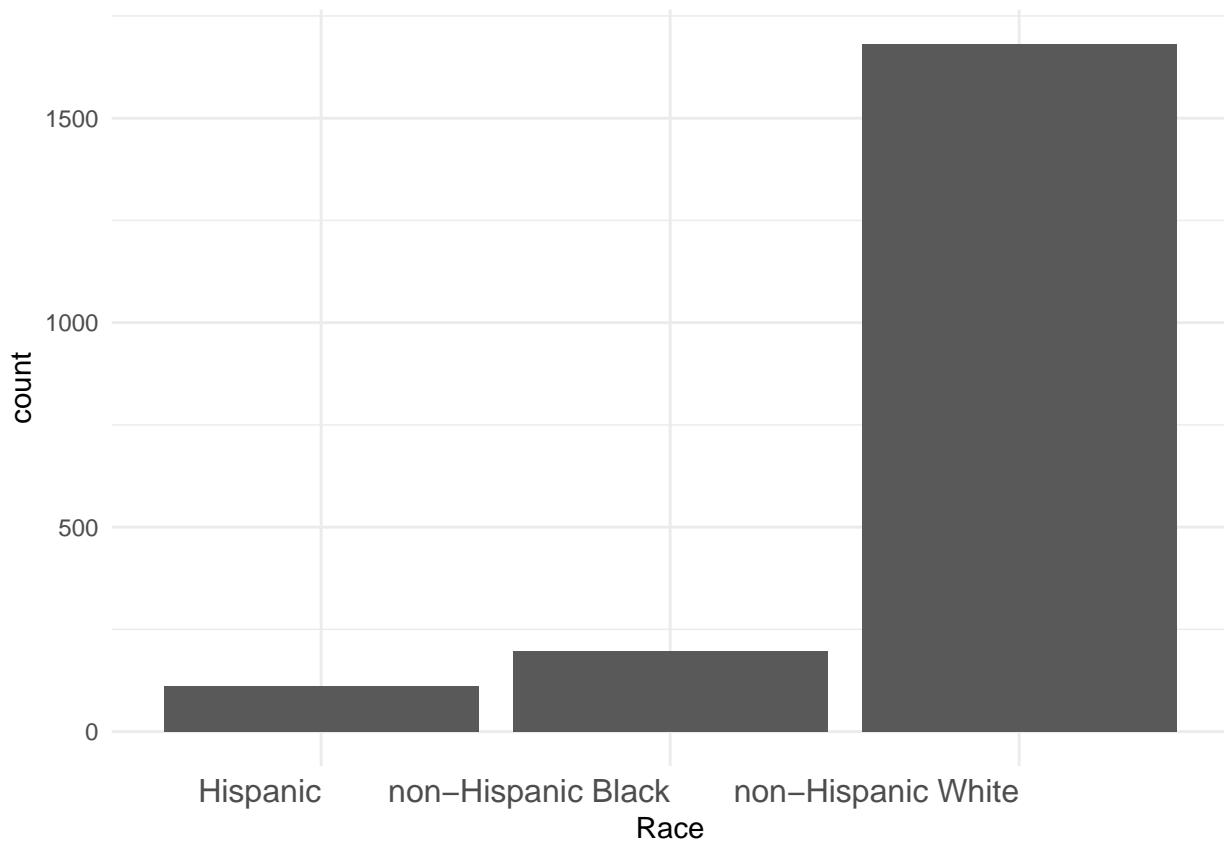
There are 15 records are missing gender (MALE) field.

ID	MARRIED	GDLIN	OBRAT	BLACK	HISPAN	MALE	APPROVE	LOANPRC	RACE
1	No	1	34.5	No	No	Unknown	Yes	75.42373	non-Hispanic White
127	No	1	31.6	No	No	Unknown	Yes	80.80000	non-Hispanic White
286	Yes	1	37.3	No	No	Unknown	Yes	80.05337	non-Hispanic White
452	Yes	1	40.1	Yes	No	Unknown	Yes	75.00000	non-Hispanic Black
618	Yes	1	38.5	No	No	Unknown	Yes	92.90323	non-Hispanic White
695	Yes	1	25.0	No	No	Unknown	Yes	64.48276	non-Hispanic White
762	Yes	1	27.6	No	No	Unknown	Yes	75.55556	non-Hispanic White
768	No	1	35.6	No	No	Unknown	Yes	64.74397	non-Hispanic White
833	Yes	1	24.0	No	Yes	Unknown	Yes	79.80769	Hispanic
979	No	1	31.7	No	No	Unknown	Yes	74.86033	non-Hispanic White
1040	No	1	38.0	Yes	No	Unknown	Yes	75.38462	non-Hispanic Black
1070	Yes	1	40.3	No	Yes	Unknown	Yes	39.39394	Hispanic
1092	Yes	1	29.7	No	No	Unknown	Yes	90.10239	non-Hispanic White
1613	No	1	30.2	No	No	Unknown	Yes	90.00000	non-Hispanic White
1924	No	1	35.8	No	No	Unknown	Yes	75.32051	non-Hispanic White

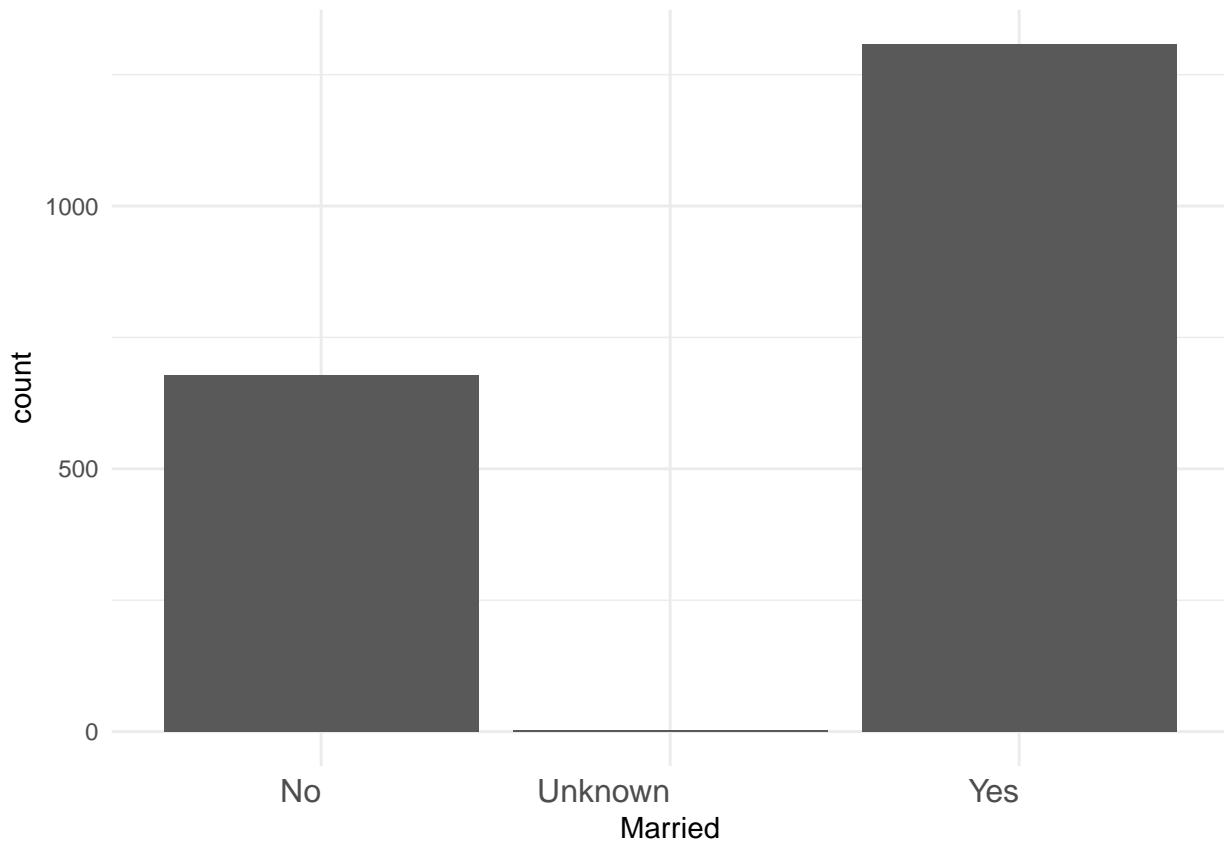
```

##          APPROVE
## RACE           No  Yes
## Hispanic        26   85
## non-Hispanic Black 64 133
## non-Hispanic White 154 1527

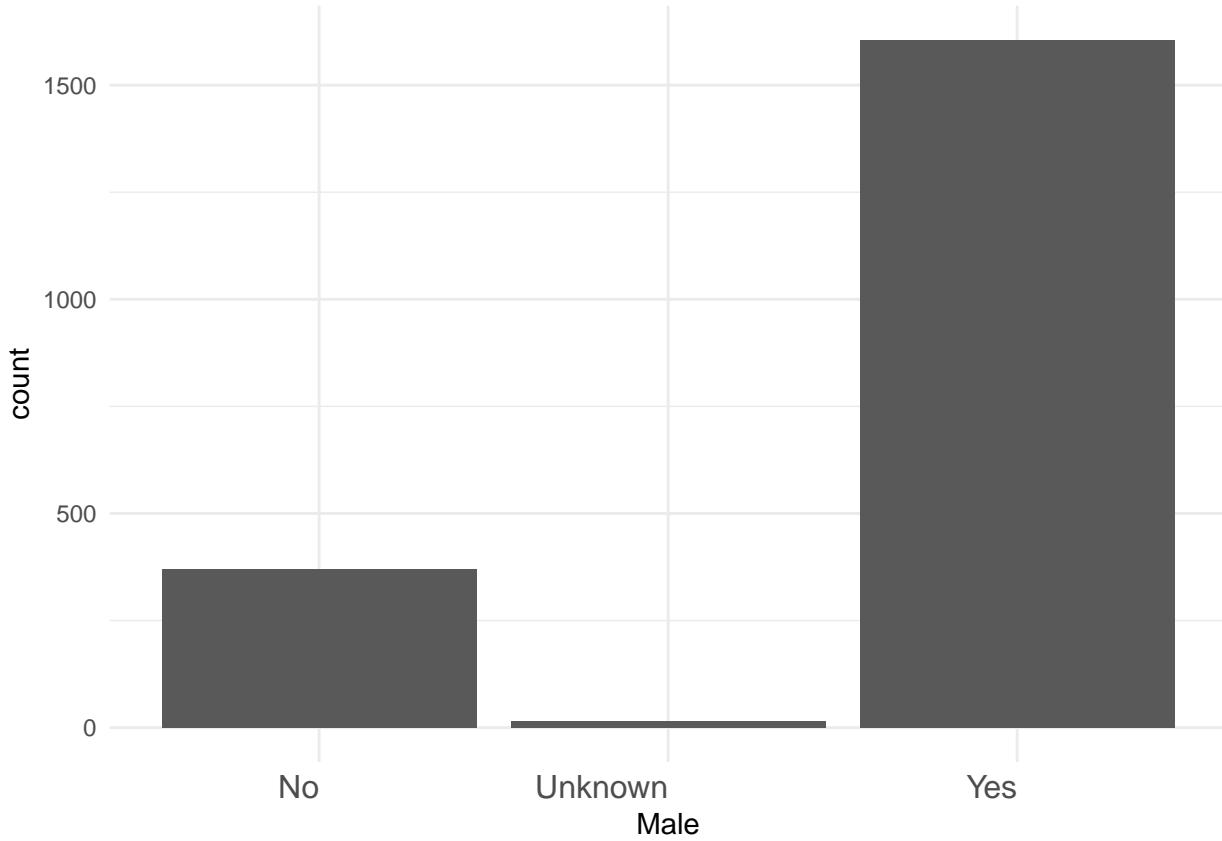
```



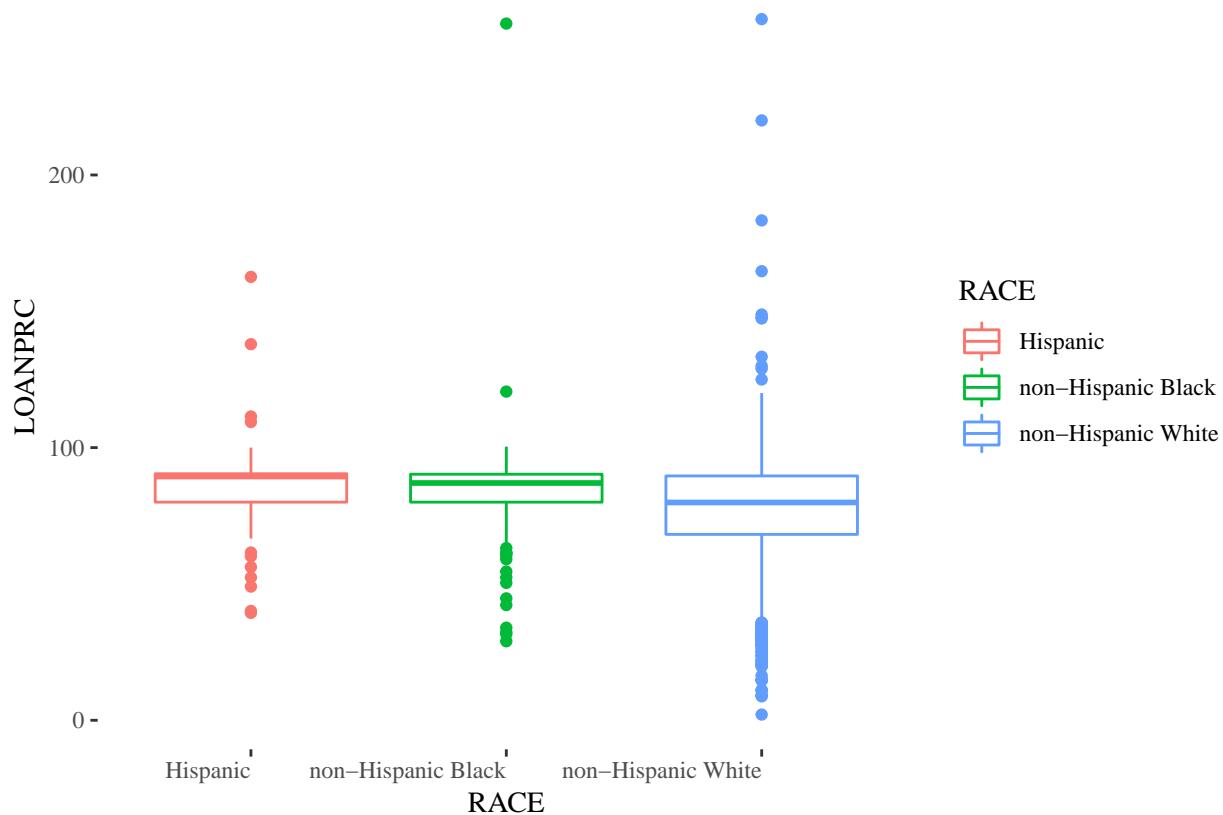
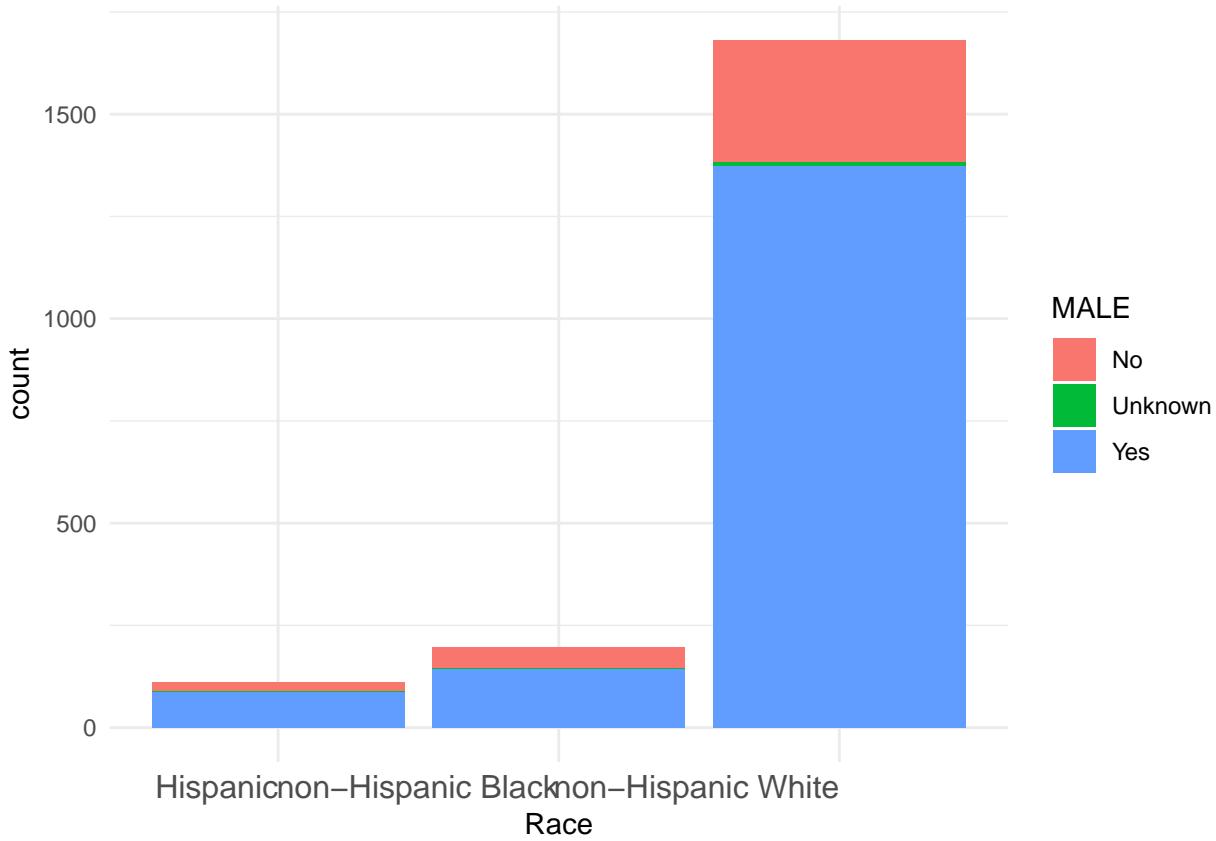
```
##          APPROVE
## MARRIED      No    Yes
##   No        102   576
## Unknown      0     3
##   Yes       142  1166
```

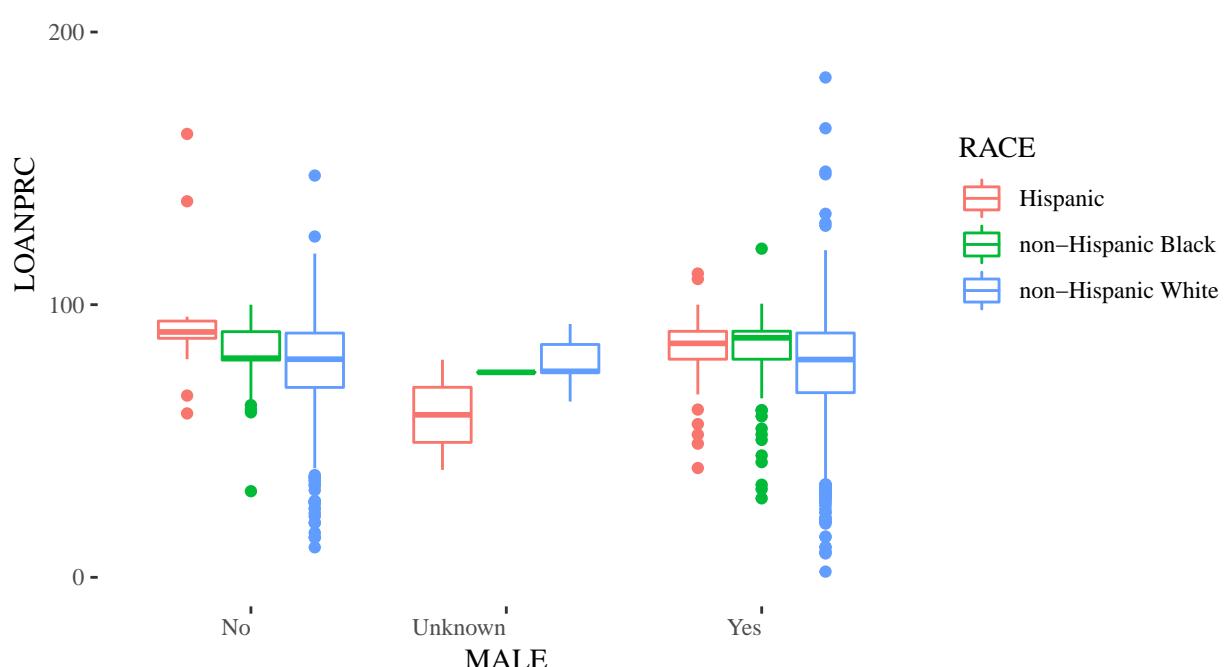
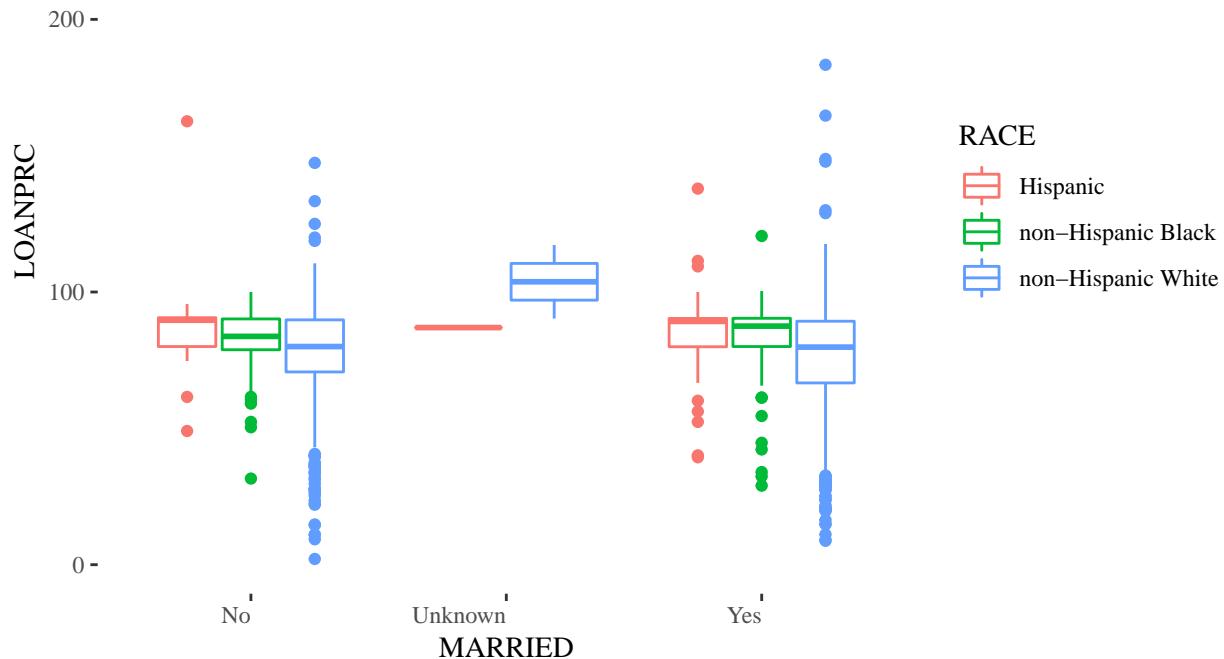


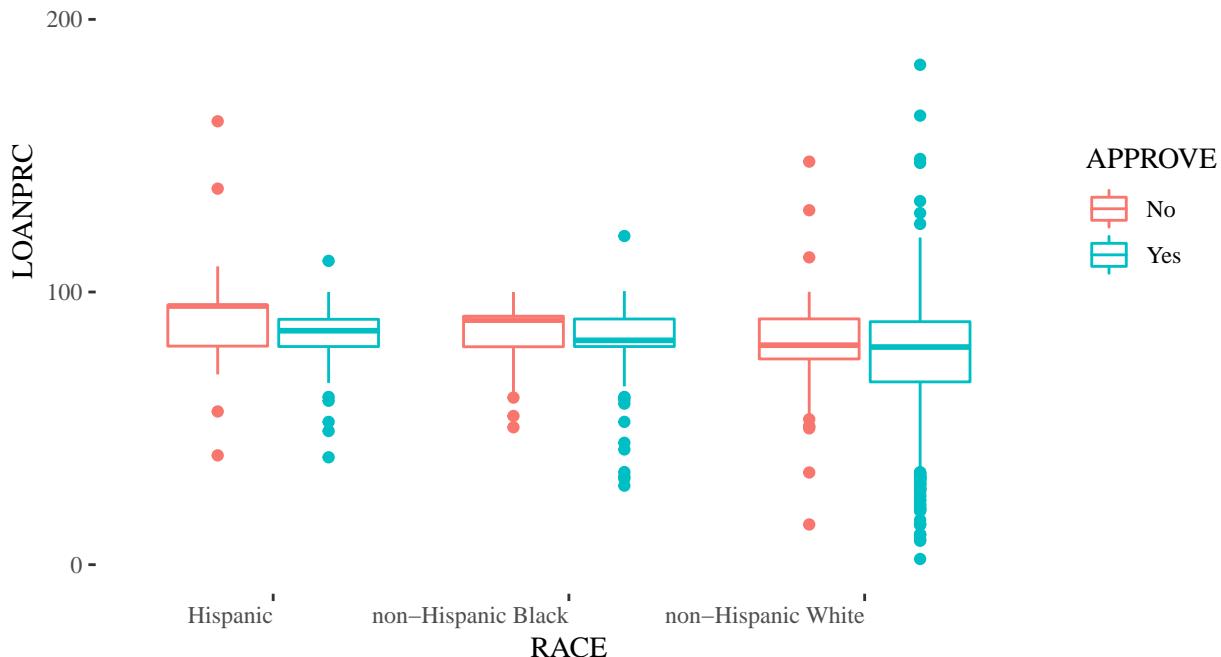
```
##          APPROVE
## MALE      No  Yes
##   No      50  319
## Unknown    0   15
##   Yes     194 1411
```



```
## , , RACE = Hispanic
##
##          APPROVE
## MARRIED      No   Yes
##   No        7    24
##   Unknown    0     1
##   Yes       19    60
##
## , , RACE = non-Hispanic Black
##
##          APPROVE
## MARRIED      No   Yes
##   No        27    49
##   Unknown    0     0
##   Yes       37    84
##
## , , RACE = non-Hispanic White
##
##          APPROVE
## MARRIED      No   Yes
##   No        68   503
##   Unknown    0     2
##   Yes       86  1022
```







```
##
## Call: glm(formula = APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE +
##           LOANPRC, family = "binomial", data = data)
##
## Coefficients:
## (Intercept)      GDLIN1       OBRAT      BLACK1      HISPAN1
##     1.53721     3.70447    -0.03525    -0.80680    -0.84499
##   MALE1      LOANPRC
##     0.14139    -0.01643
##
## Degrees of Freedom: 1971 Total (i.e. Null); 1965 Residual
## Null Deviance: 1476
## Residual Deviance: 966 AIC: 980
##
##              2.5 %      97.5 %
## (Intercept) 0.39043496 2.703252996
## GDLIN1      3.29115094 4.137733306
## OBRAT      -0.05549706 -0.015038547
## BLACK1      -1.26962817 -0.326297080
## HISPAN1     -1.42913599 -0.215005260
## MALE1       -0.30298683  0.565437899
## LOANPRC     -0.02663168 -0.006658542
##
##              2.5 %      97.5 %
## (Intercept) 0.38363811 2.690789543
## GDLIN1      3.28199418 4.126951904
## OBRAT      -0.05547873 -0.015022012
## BLACK1      -1.27779072 -0.335818281
```

```

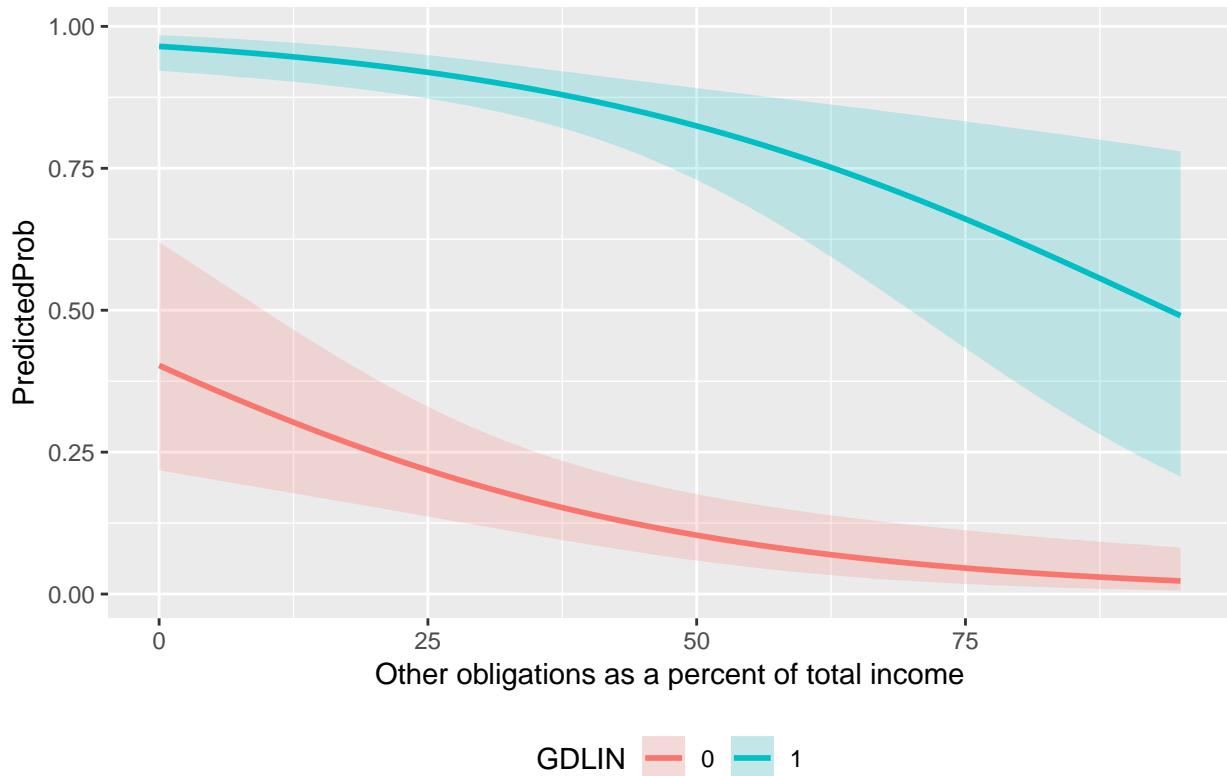
## HISPAN1      -1.45043209 -0.239548275
## MALE1        -0.29193829  0.574720492
## LOANPRC      -0.02635892 -0.006504892

## (Intercept)    GDLIN1      OBRAT      BLACK1      HISPAN1      MALE1
## 4.6516120   40.6286322   0.9653637   0.4462819   0.4295616   1.1518751
## LOANPRC
## 0.9837024

##          OR      2.5 %     97.5 %
## (Intercept) 4.6516120 1.4776234 14.9282142
## GDLIN1      40.6286322 26.8737761 62.6606279
## OBRAT       0.9653637  0.9460148  0.9850740
## BLACK1       0.4462819  0.2809361  0.7215908
## HISPAN1     0.4295616  0.2395158  0.8065372
## MALE1        1.1518751  0.7386088  1.7602184
## LOANPRC     0.9837024  0.9737198  0.9933636

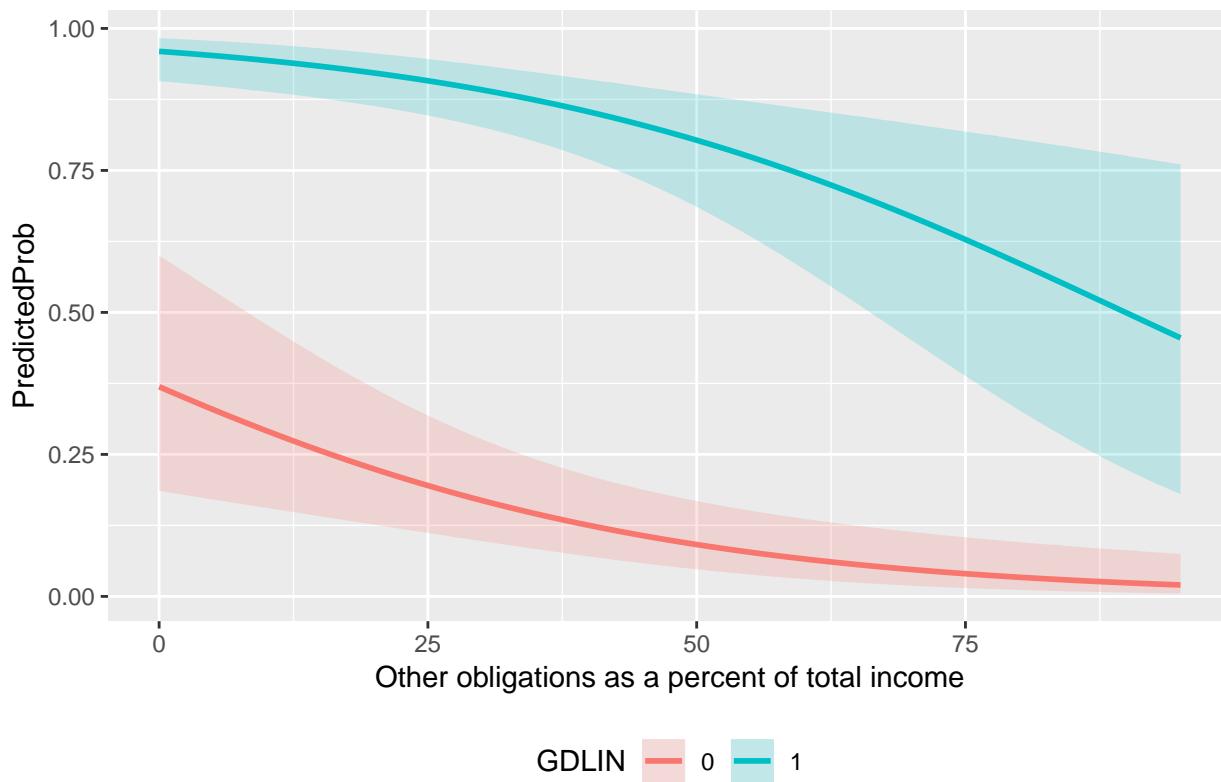
```

non-Hispanic Black Male Predicted probabilities

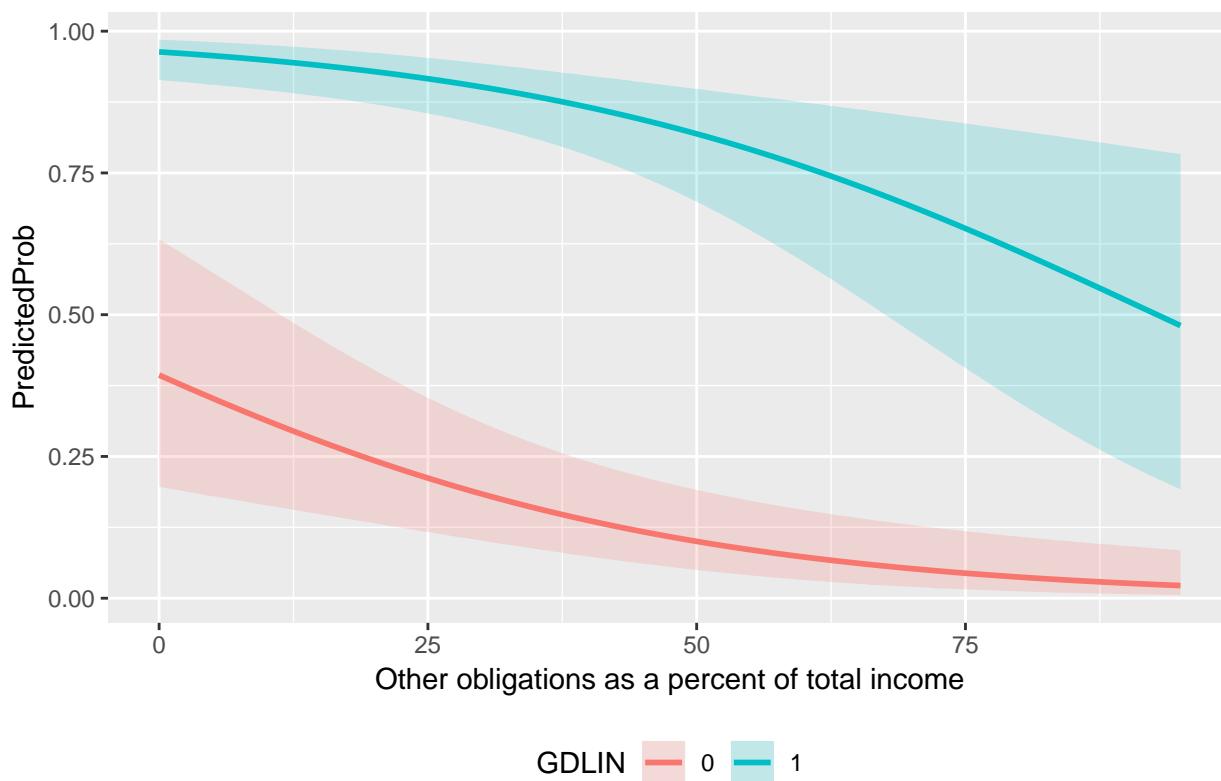


GDLIN 0 1

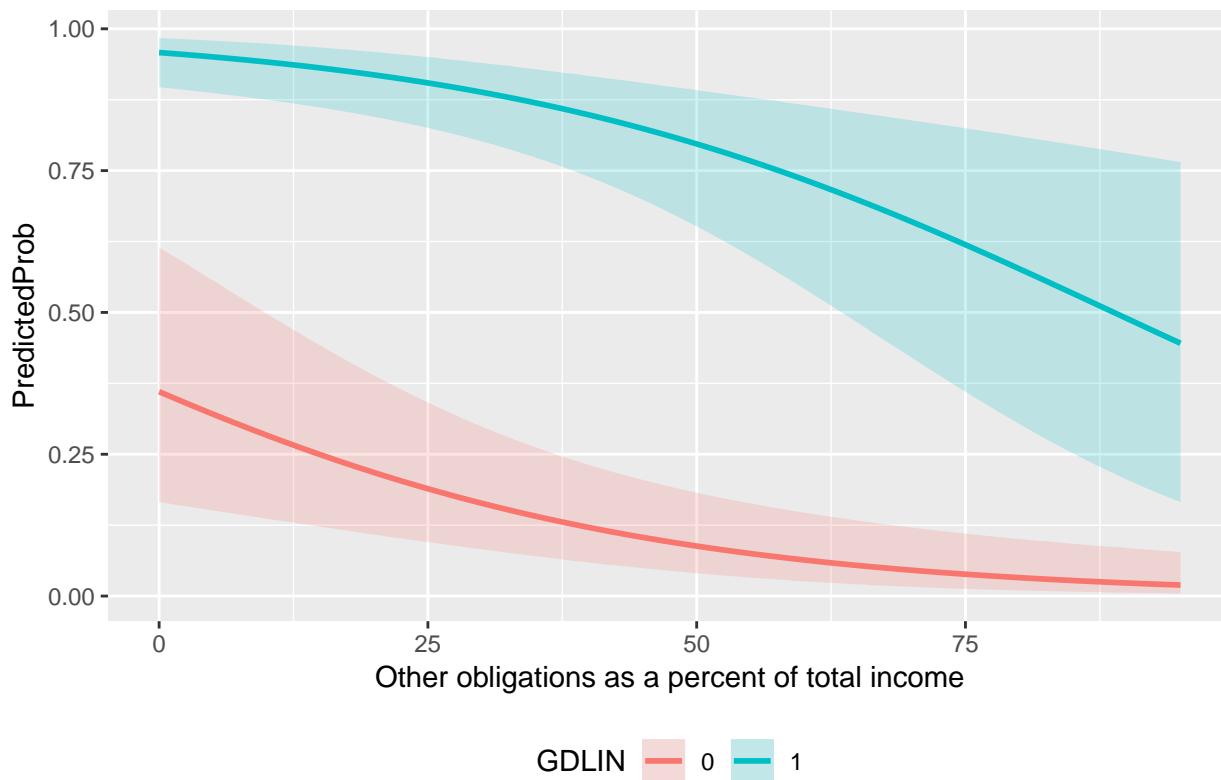
non-Hispanic Black Female Predicted probabilities



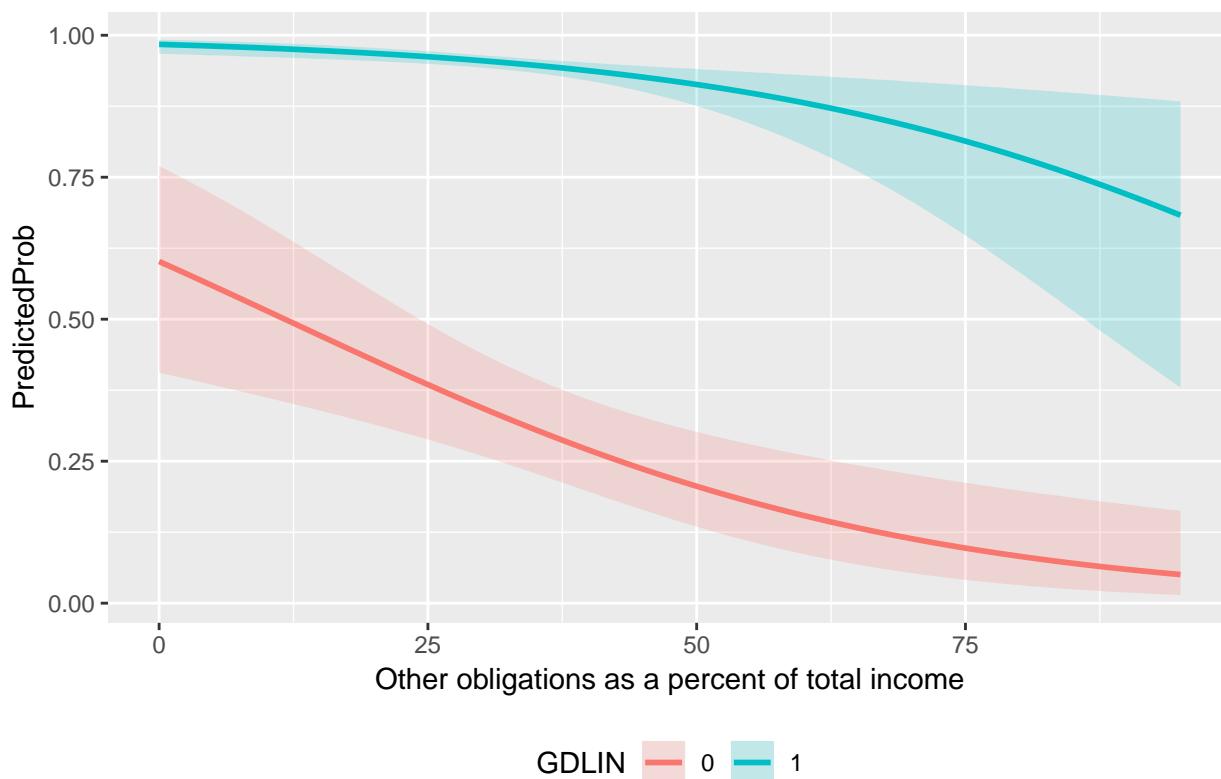
Hispanic Male Predicted probabilities



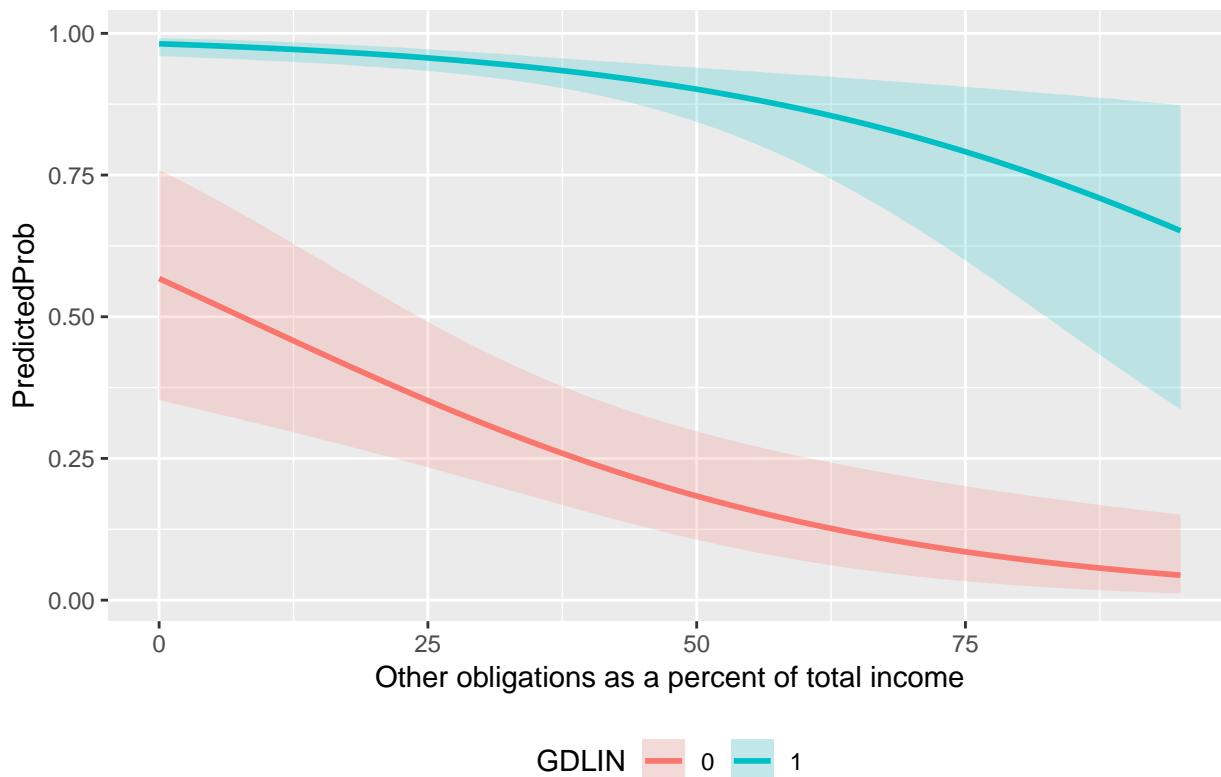
Hispanic Female Predicted probabilities



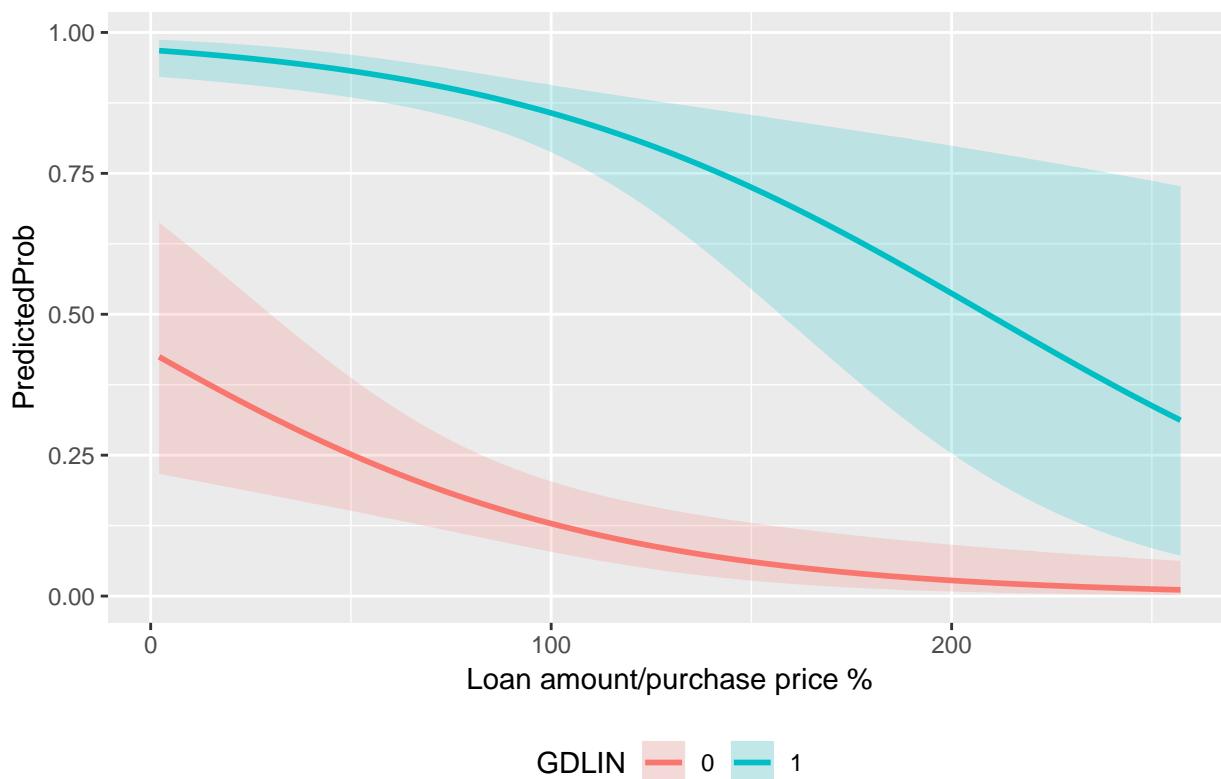
non-Hispanic White Male Predicted probabilities



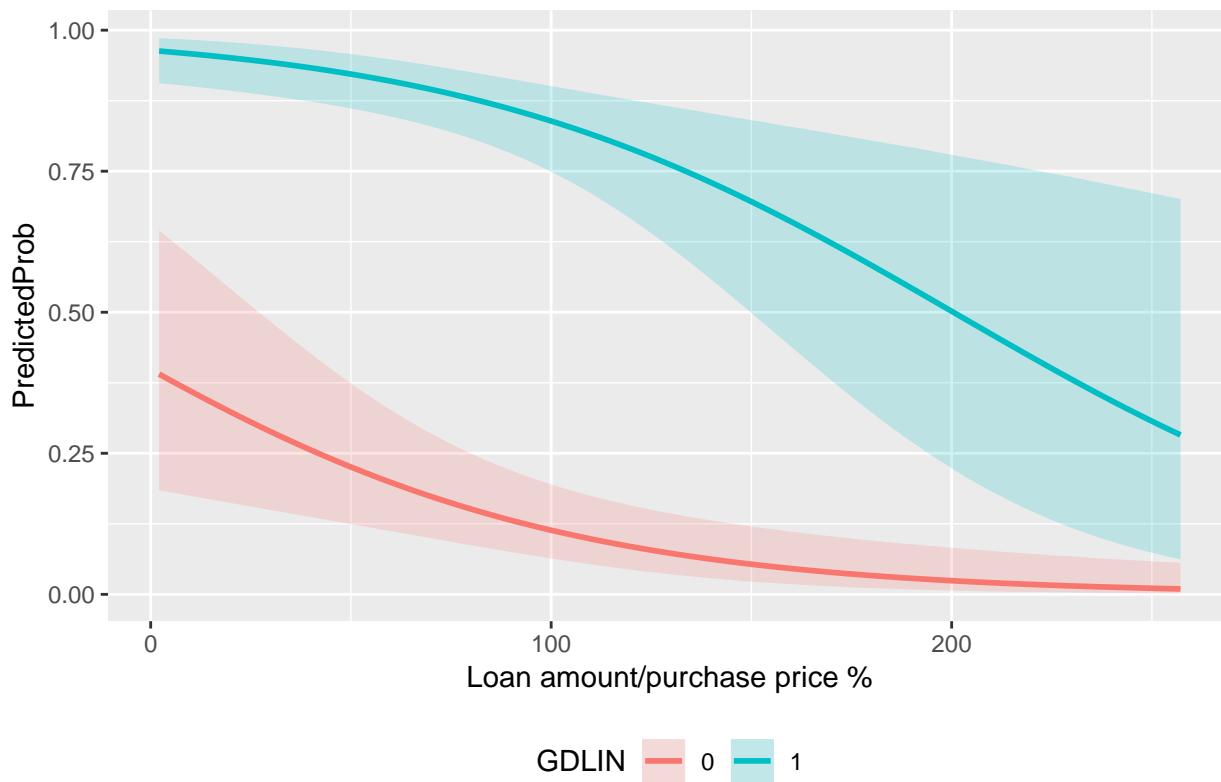
non-Hispanic White Female Predicted probabilities



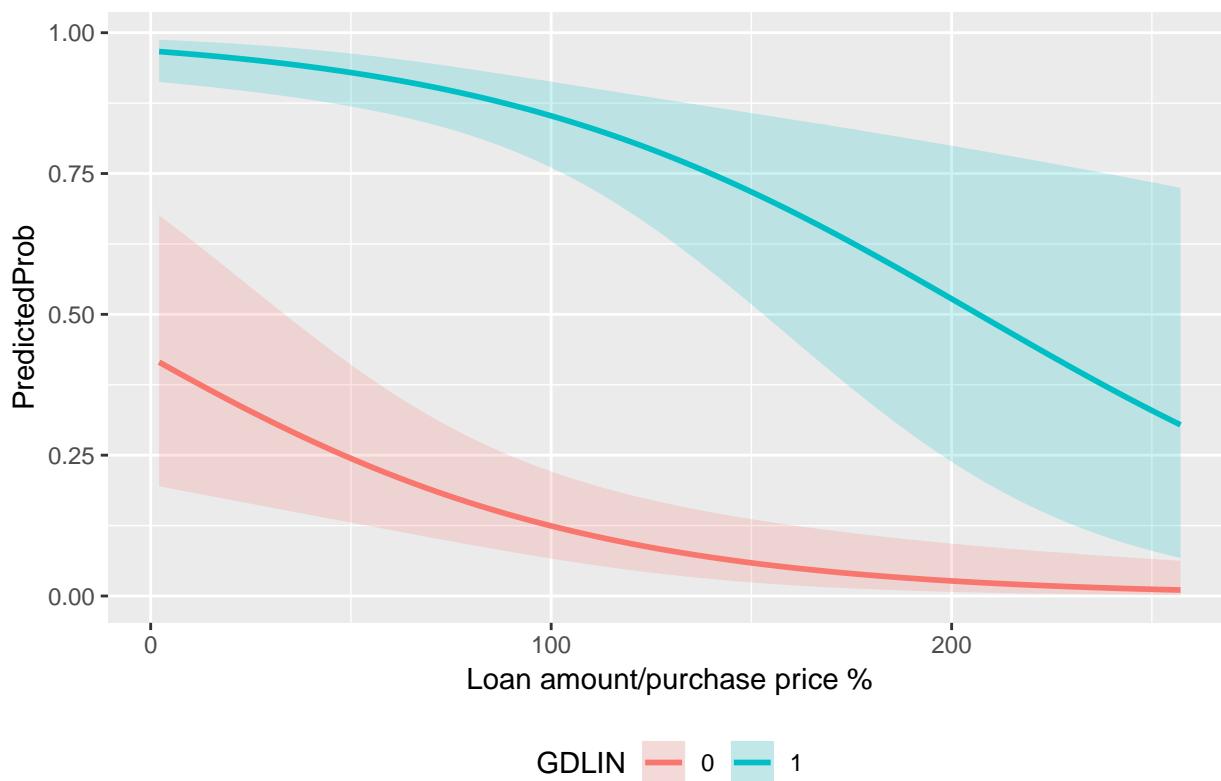
non-Hispanic Black Male Predicted probabilities



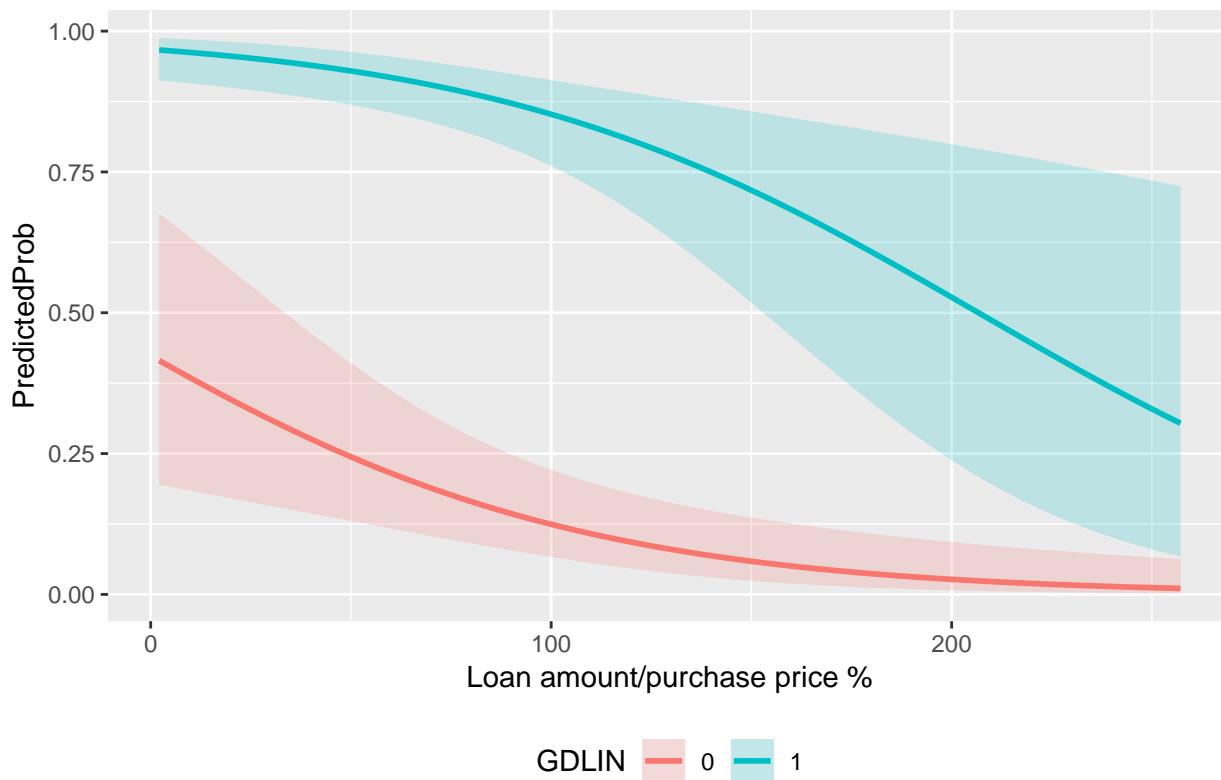
non-Hispanic Black Female Predicted probabilities



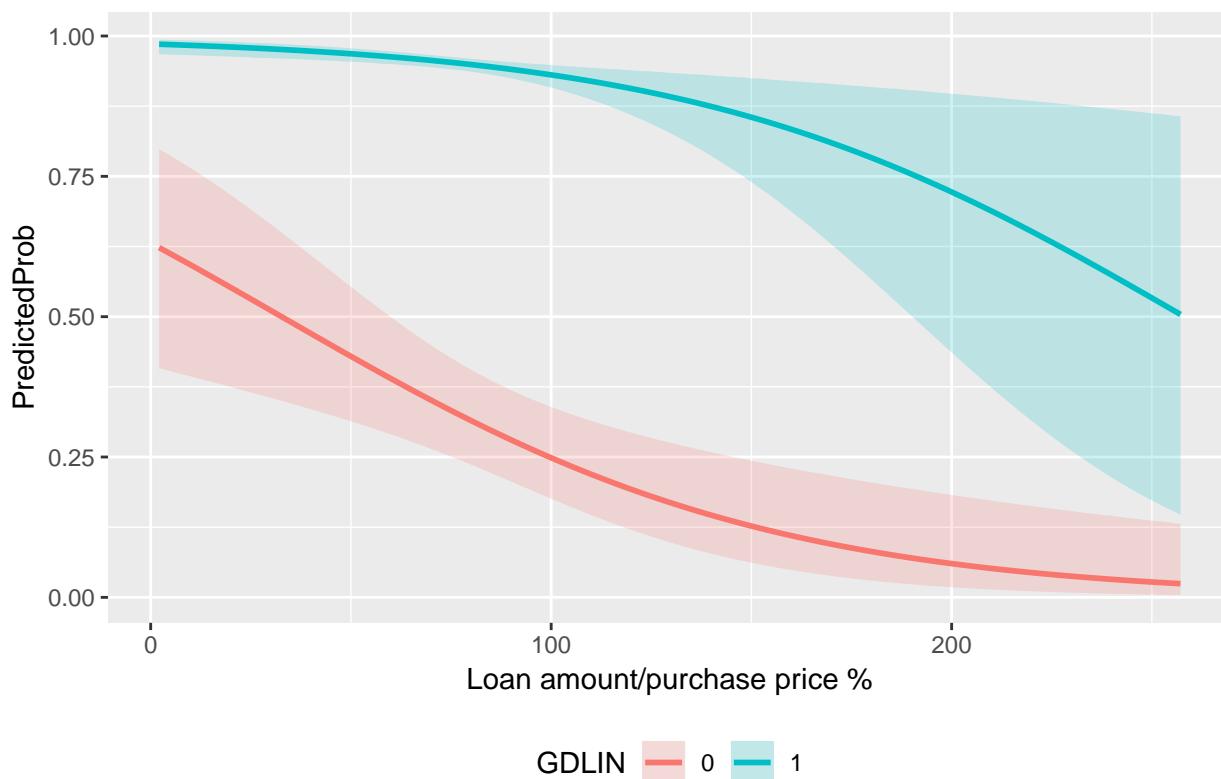
Hispanic Male Predicted probabilities



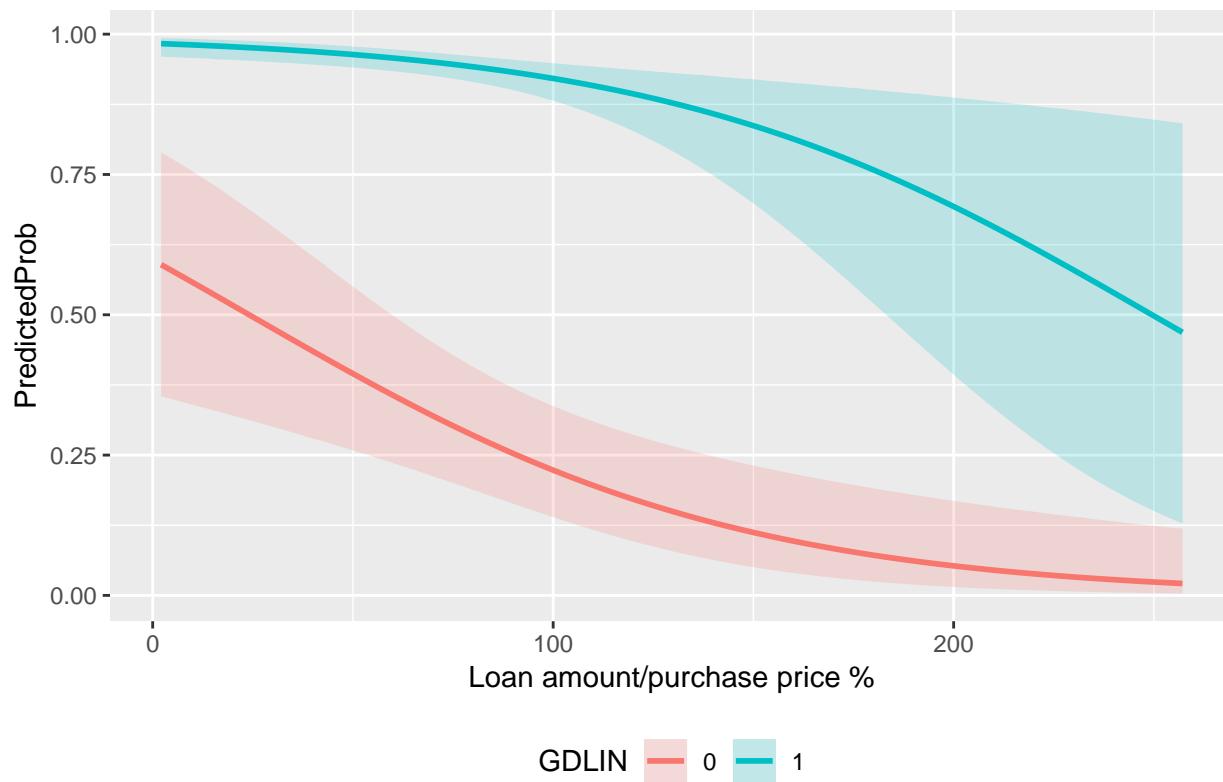
Hispanic Female Predicted probabilities



non-Hispanic White Male Predicted probabilities



non-Hispanic White Female Predicted probabilities



```

## 
## Call: glm(formula = APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + LOANPRC,
##           family = "binomial", data = data)
## 
## Coefficients:
## (Intercept)      GDLIN1       OBRAT       BLACK1      HISPAN1
## 1.66644        3.69659     -0.03524     -0.81891     -0.84836
## LOANPRC
## -0.01652
## 
## Degrees of Freedom: 1971 Total (i.e. Null); 1966 Residual
## Null Deviance: 1476
## Residual Deviance: 966.4      AIC: 978.4

## 
## Call: glm(formula = APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE +
##           LOANPRC, family = binomial(link = "probit"), data = data)
## 
## Coefficients:
## (Intercept)      GDLIN1       OBRAT       BLACK1      HISPAN1
## 0.64766        2.13928     -0.01705     -0.42091     -0.43124
## MALE1          LOANPRC
## 0.06307       -0.00827
## 
## Degrees of Freedom: 1971 Total (i.e. Null); 1965 Residual
## Null Deviance: 1476
## Residual Deviance: 965.5      AIC: 979.5

## Overall

```

```

## GDLIN1 17.7034125
## OBRAT   3.1949750
## BLACK1   3.3211862
## HISPAN1  2.6512623
## MALE1    0.5695835
## LOANPRC  3.2178042

##
## Call: glm(formula = APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + LOANPRC,
##           family = binomial(link = "probit"), data = data)
##
## Coefficients:
## (Intercept)      GDLIN1        OBRAT        BLACK1        HISPAN1
## 0.706553     2.136046     -0.017060     -0.427179     -0.432845
## LOANPRC
## -0.008317
##
## Degrees of Freedom: 1971 Total (i.e. Null); 1966 Residual
## Null Deviance: 1476
## Residual Deviance: 965.8      AIC: 977.8

## Overall
## GDLIN1 17.702452
## OBRAT   3.197948
## BLACK1   3.383556
## HISPAN1  2.662084
## LOANPRC  3.236990

## Analysis of Deviance Table
##
## Model 1: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE + LOANPRC
## Model 2: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + LOANPRC
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       1965     965.48
## 2       1966     965.80 -1   -0.3199    0.5717

## Likelihood ratio test
##
## Model 1: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE + LOANPRC
## Model 2: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + LOANPRC
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 7 -482.74
## 2 6 -482.90 -1 0.3199    0.5717

##
## Call: glm(formula = APPROVE ~ GDLIN + BLACK + HISPAN + MALE + LOANPRC,
##           family = binomial(link = "probit"), data = data)
##
## Coefficients:
## (Intercept)      GDLIN1        BLACK1        HISPAN1        MALE1
## 0.105974     2.184604     -0.441909     -0.441348     0.064210
## LOANPRC
## -0.009121
##
## Degrees of Freedom: 1971 Total (i.e. Null); 1966 Residual
## Null Deviance: 1476
## Residual Deviance: 977.2      AIC: 989.2

```

```

## Overall
## GDLIN1 18.2133105
## BLACK1  3.4982924
## HISPAN1 2.7352593
## MALE1   0.5837877
## LOANPRC 3.6344426

## Analysis of Deviance Table
##
## Model 1: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE + LOANPRC
## Model 2: APPROVE ~ GDLIN + BLACK + HISPAN + MALE + LOANPRC
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1965    965.48
## 2      1966    977.18 -1   -11.697 0.0006261 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test
##
## Model 1: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE + LOANPRC
## Model 2: APPROVE ~ GDLIN + BLACK + HISPAN + MALE + LOANPRC
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1    7 -482.74
## 2    6 -488.59 -1 11.697 0.0006261 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call: glm(formula = APPROVE ~ GDLIN + BLACK + HISPAN + LOANPRC, family = binomial(link = "probit"),
##           data = data)
##
## Coefficients:
## (Intercept)      GDLIN1       BLACK1       HISPAN1       LOANPRC
## 0.165274     2.181533    -0.448539    -0.443199    -0.009168
##
## Degrees of Freedom: 1971 Total (i.e. Null); 1967 Residual
## Null Deviance: 1476
## Residual Deviance: 977.5      AIC: 987.5

## Overall
## GDLIN1 18.213136
## BLACK1  3.565111
## HISPAN1 2.748035
## LOANPRC 3.654179

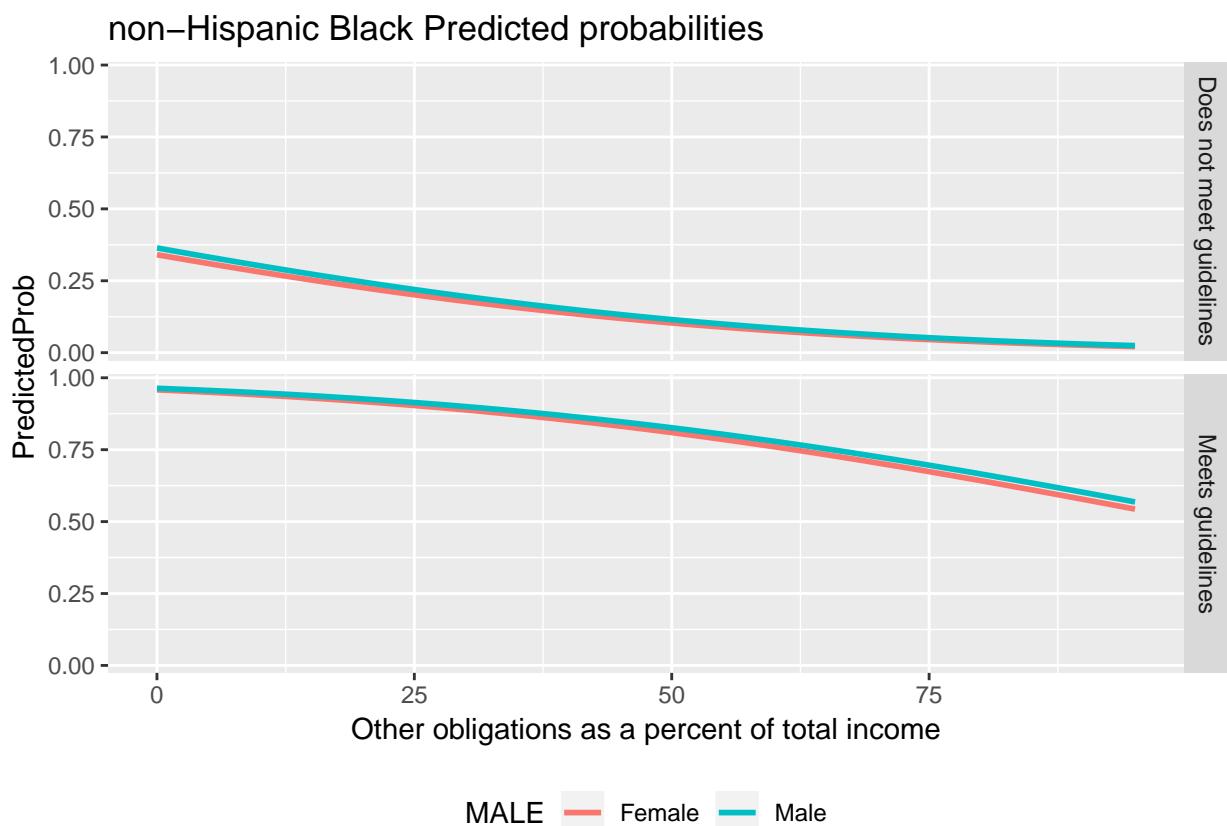
## Analysis of Deviance Table
##
## Model 1: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE + LOANPRC
## Model 2: APPROVE ~ GDLIN + BLACK + HISPAN + LOANPRC
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1965    965.48
## 2      1967    977.52 -2   -12.033 0.002438 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

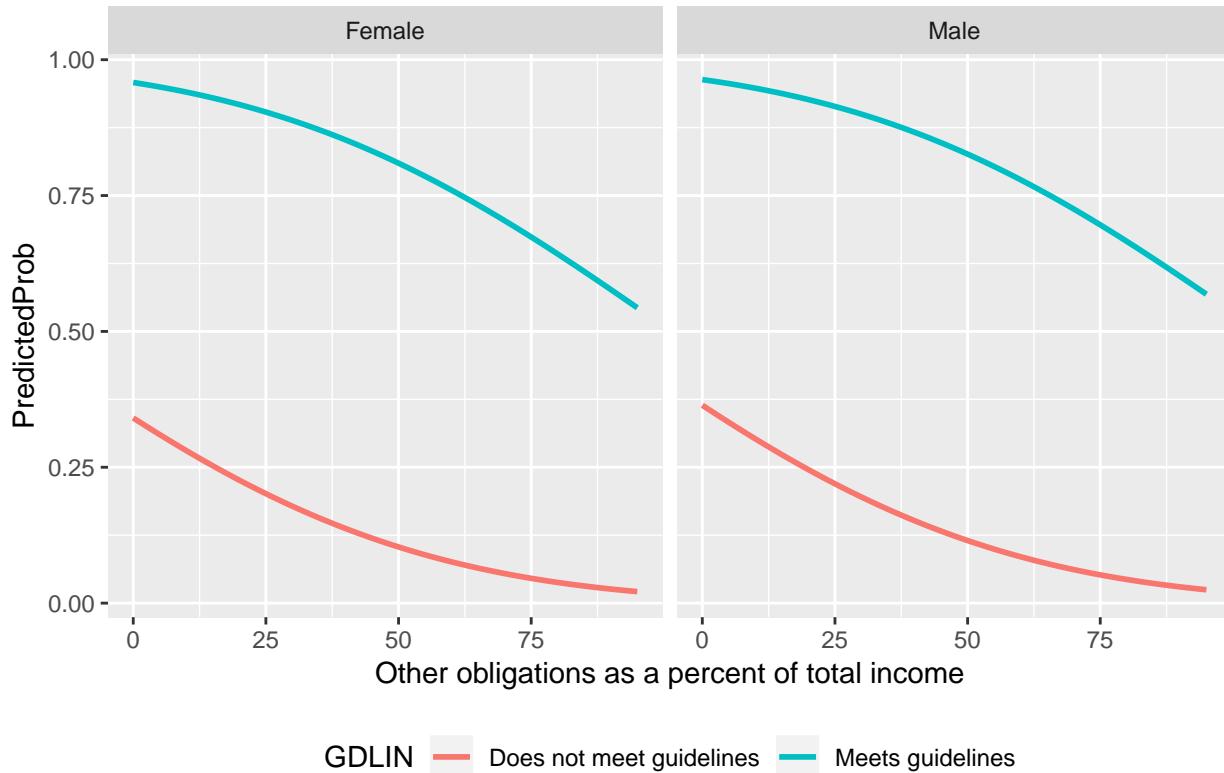
```

## Likelihood ratio test
##
## Model 1: APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + MALE + LOANPRC
## Model 2: APPROVE ~ GDLIN + BLACK + HISPAN + LOANPRC
## #Df LogLik Df Chisq Pr(>Chisq)
## 1    7 -482.74
## 2    5 -488.76 -2 12.033   0.0002438 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

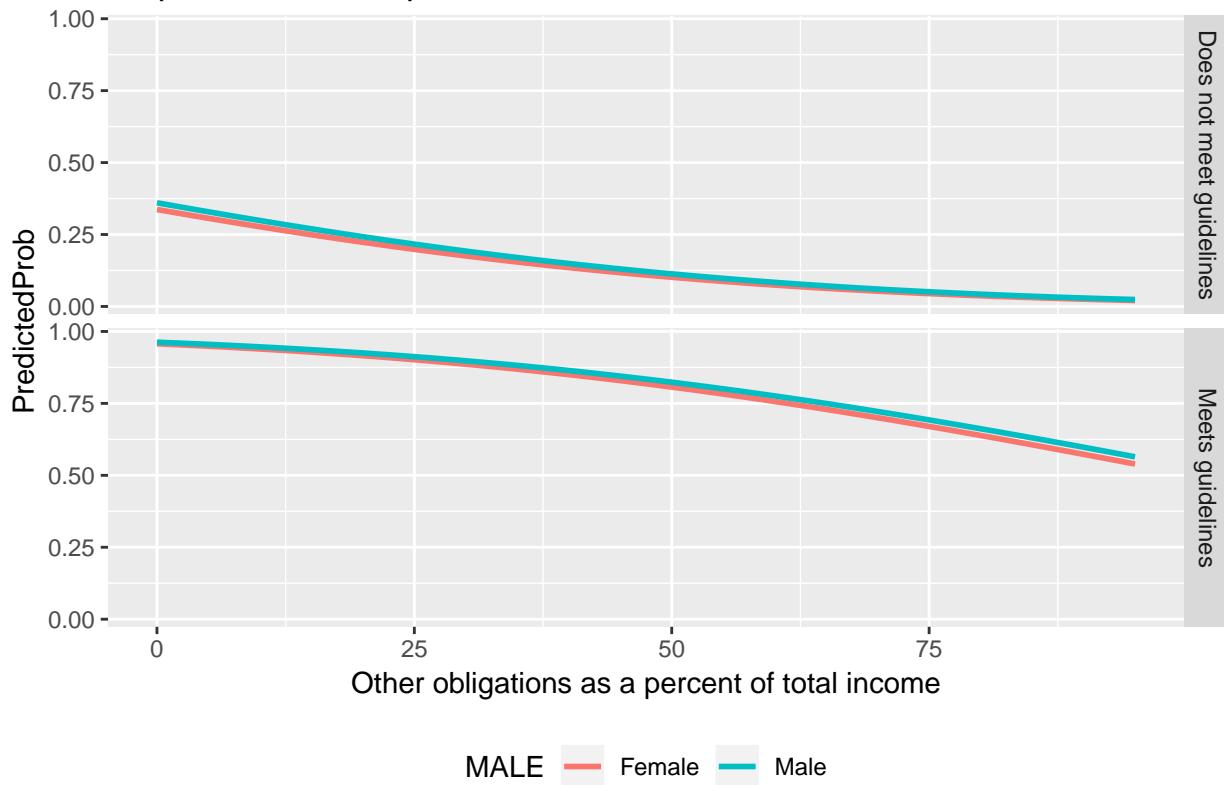
```



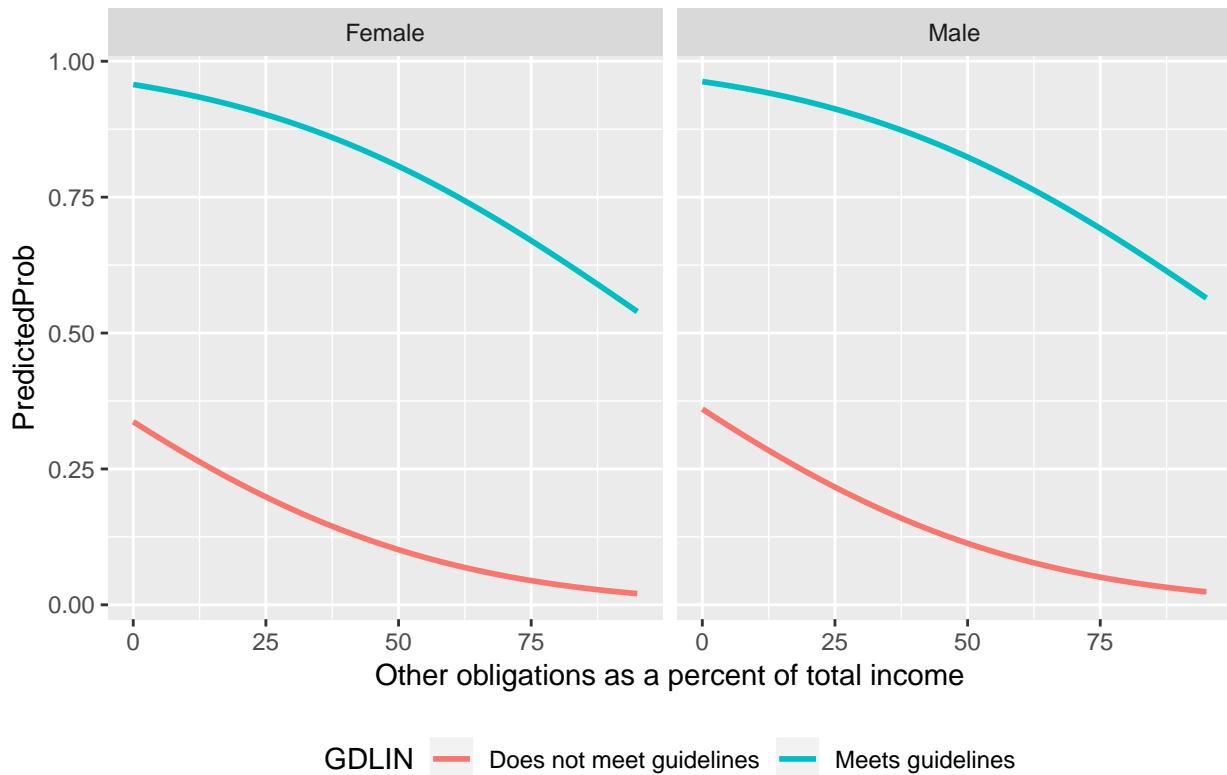
non-Hispanic Black Predicted probabilities



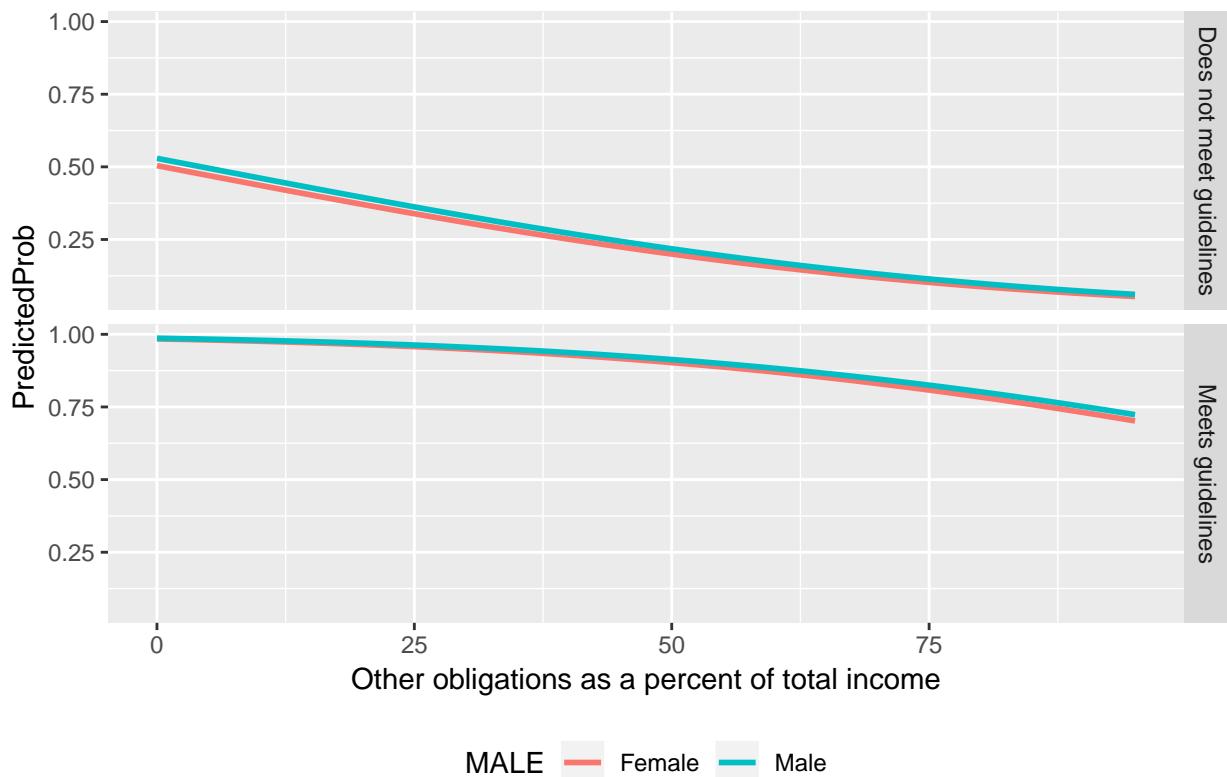
Hispanic Predicted probabilities



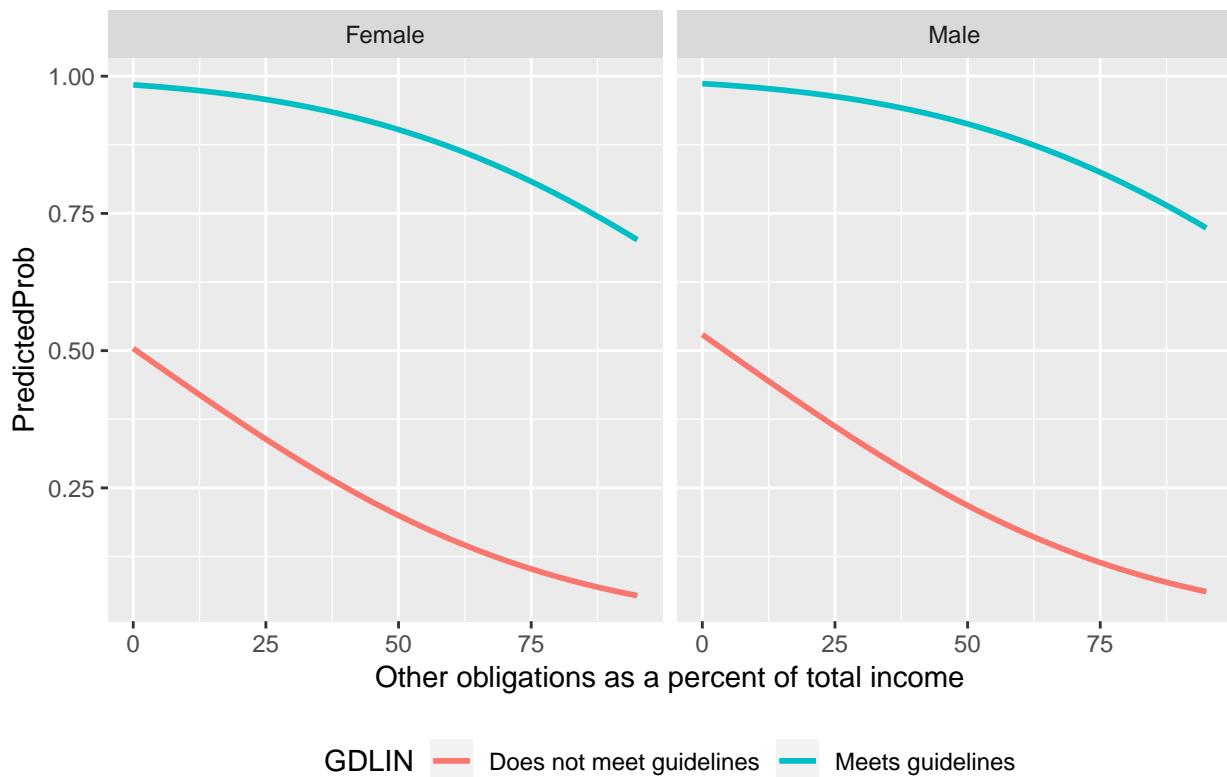
Hispanic Predicted probabilities



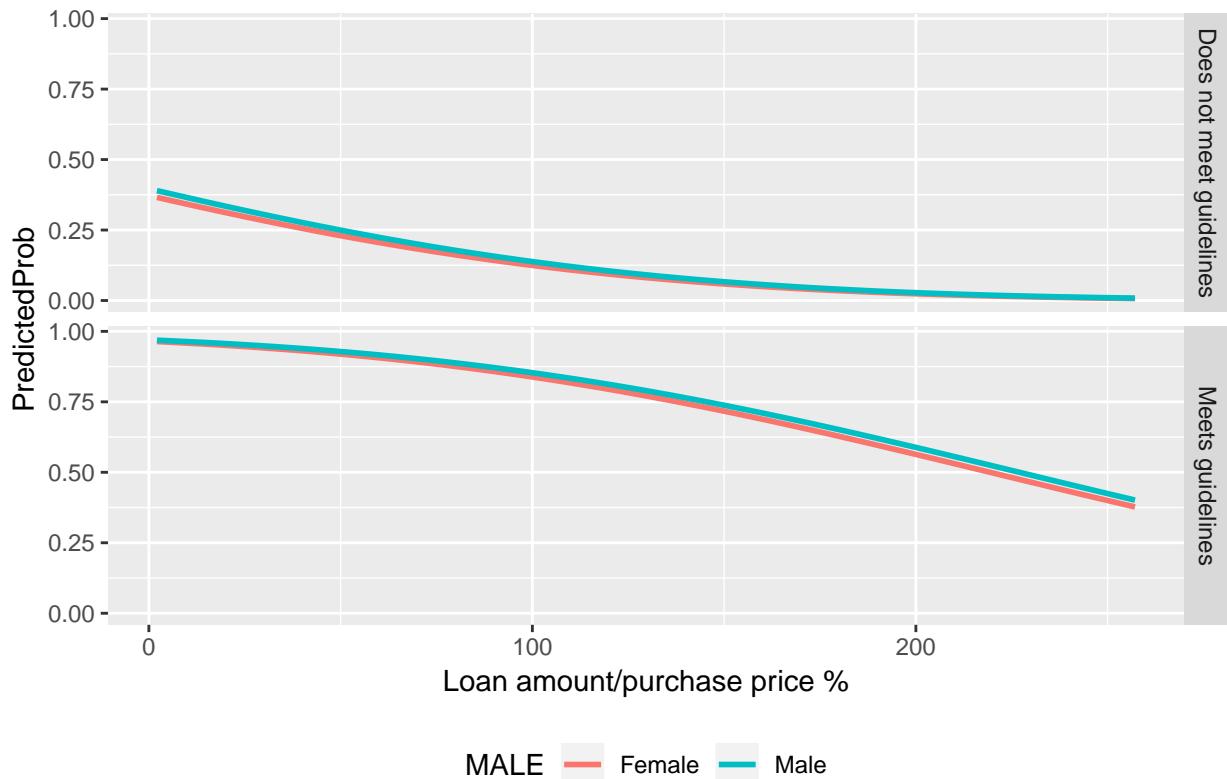
non-Hispanic White Predicted probabilities



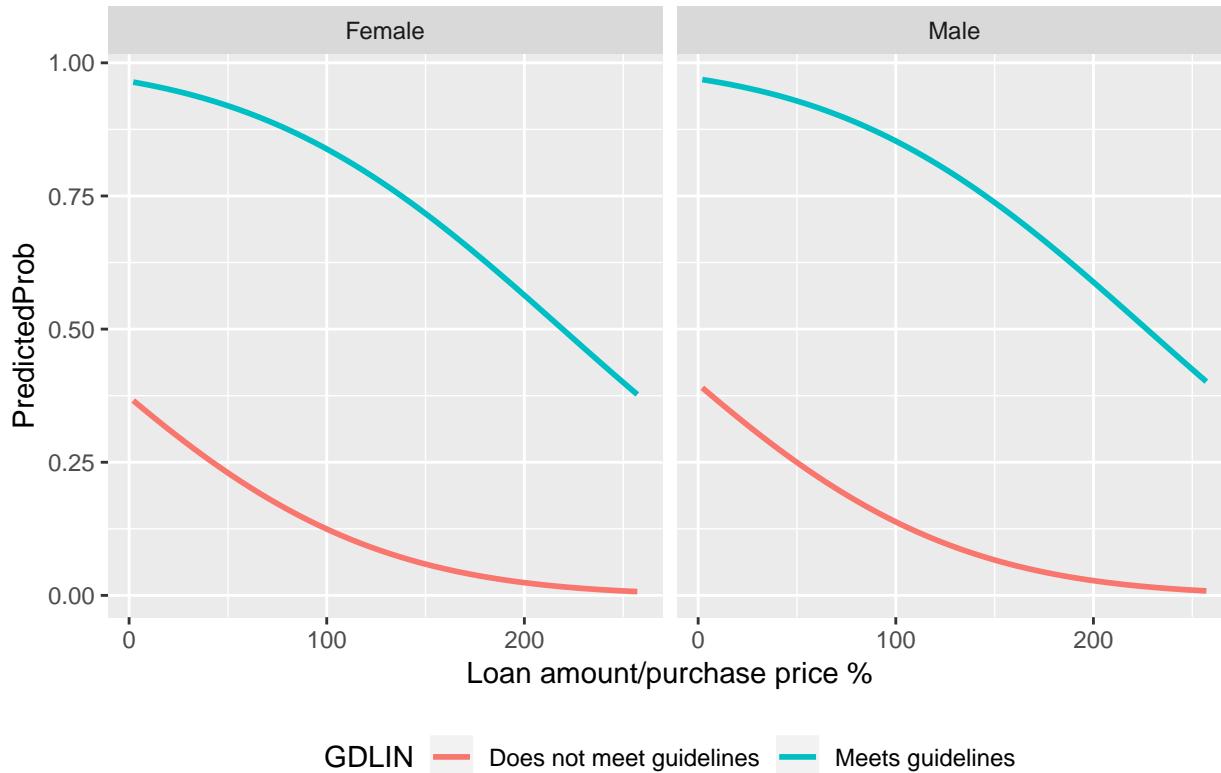
non-Hispanic White Predicted probabilities



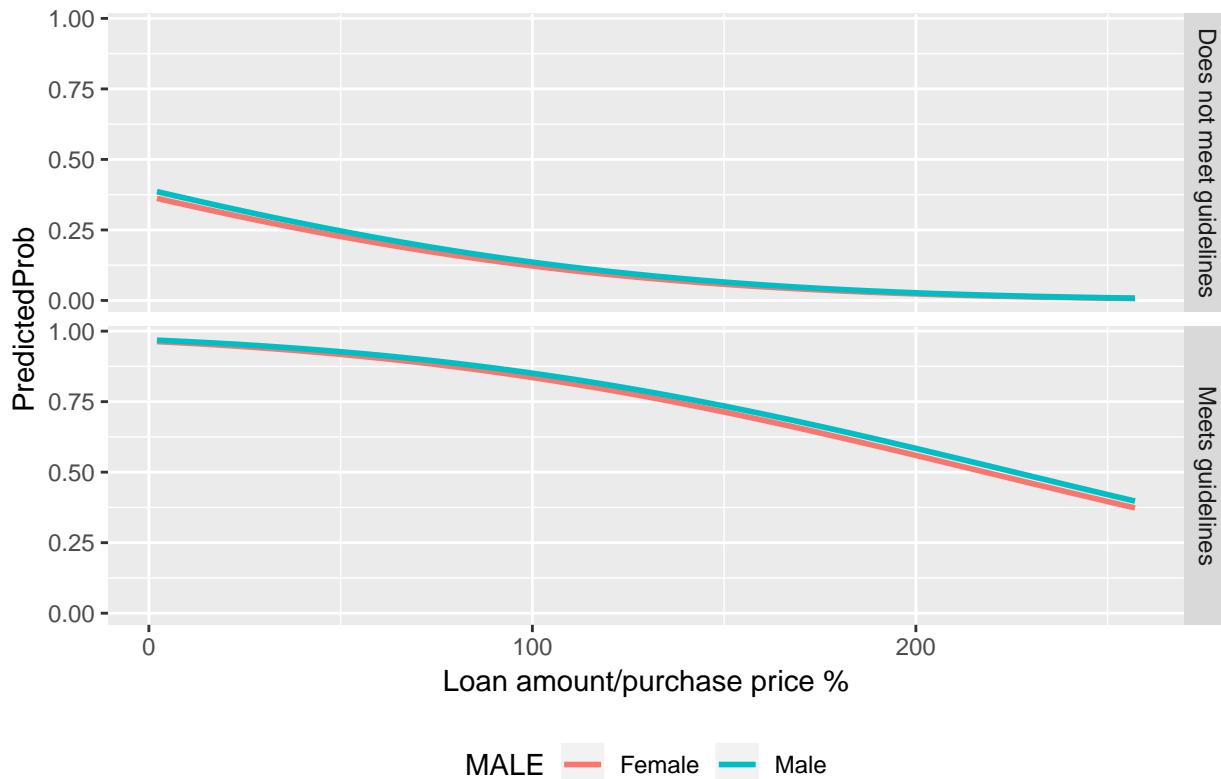
non-Hispanic Black Predicted probabilities



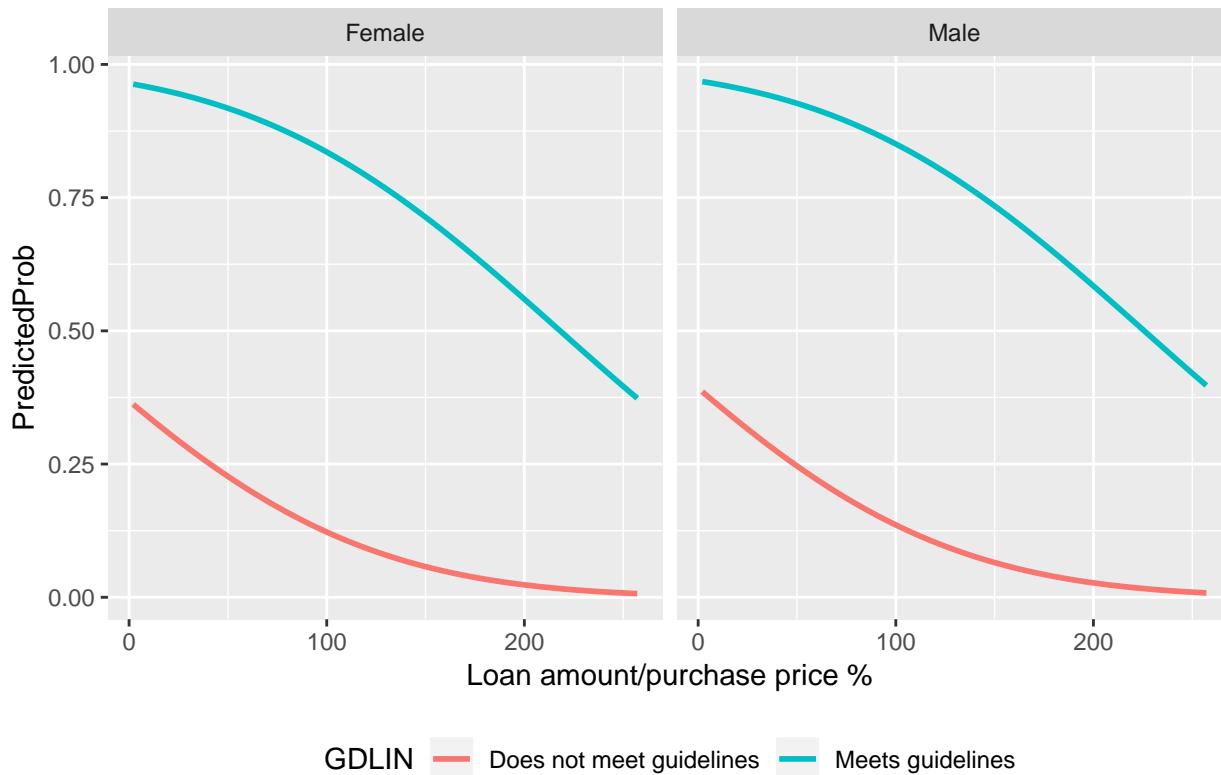
non-Hispanic Black Predicted probabilities



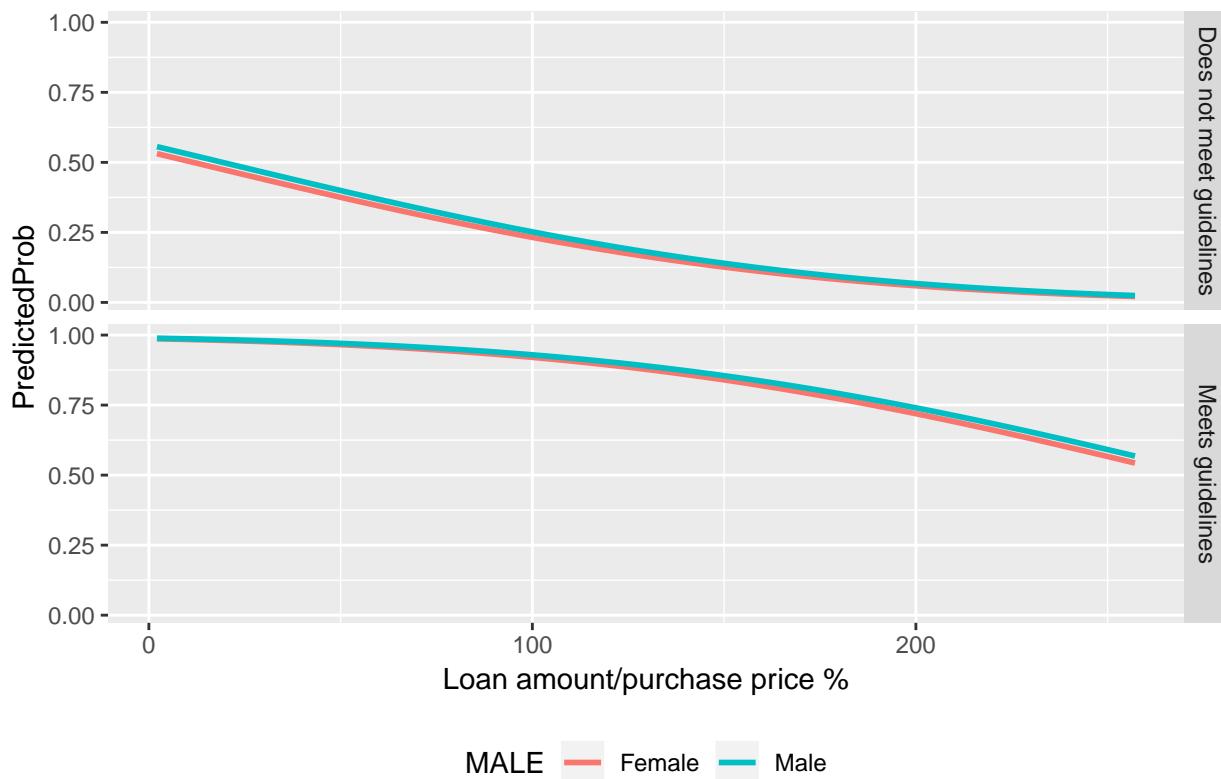
Hispanic Predicted probabilities



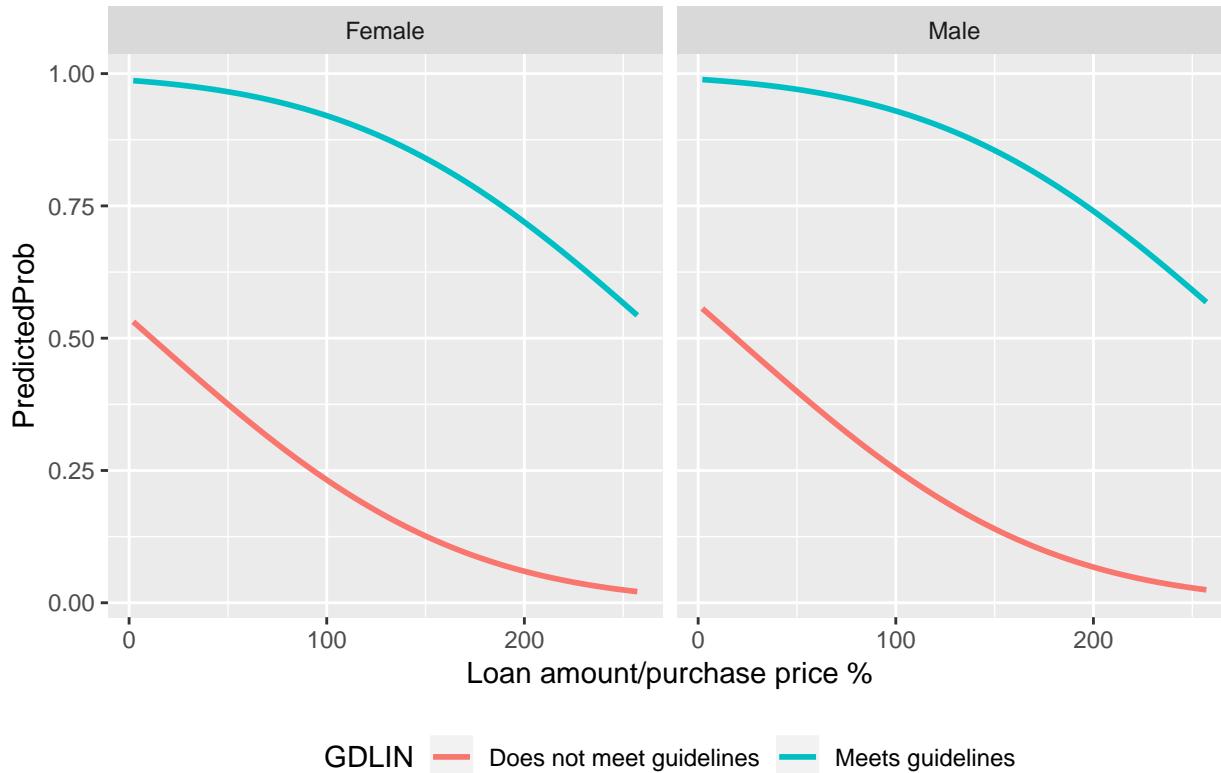
Hispanic Predicted probabilities



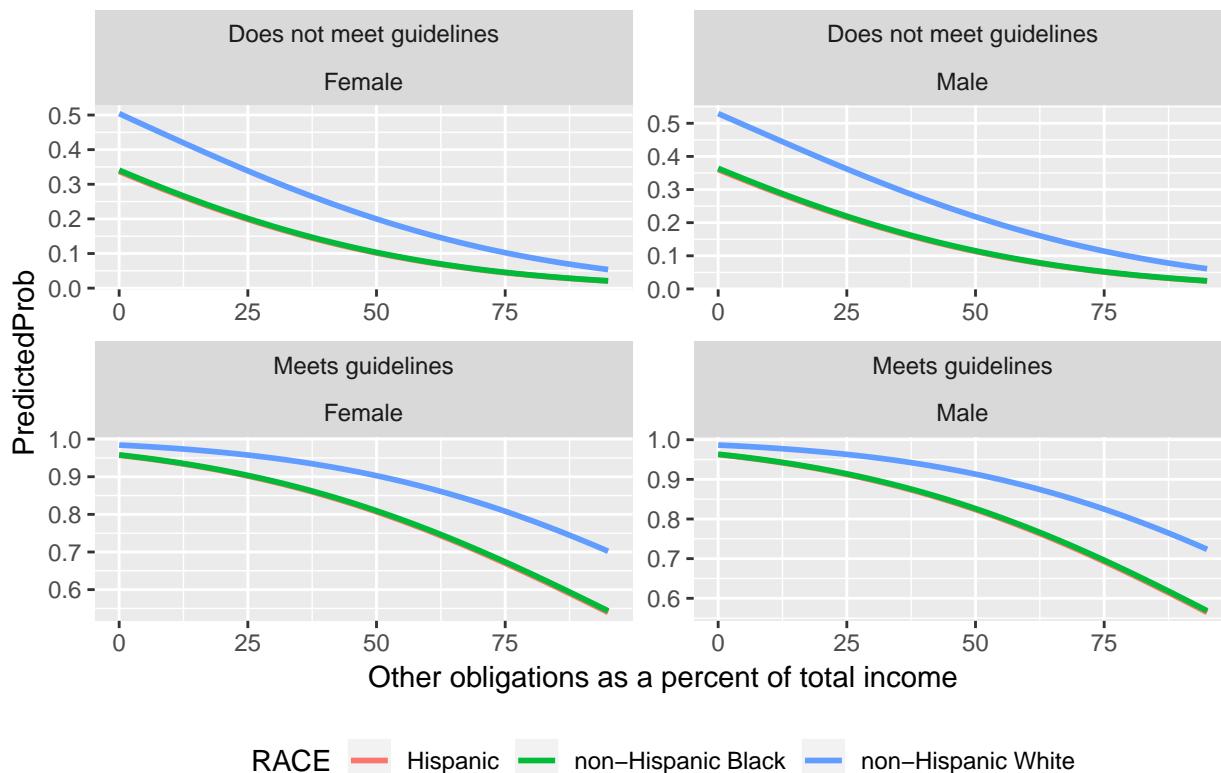
non-Hispanic White Black Predicted probabilities



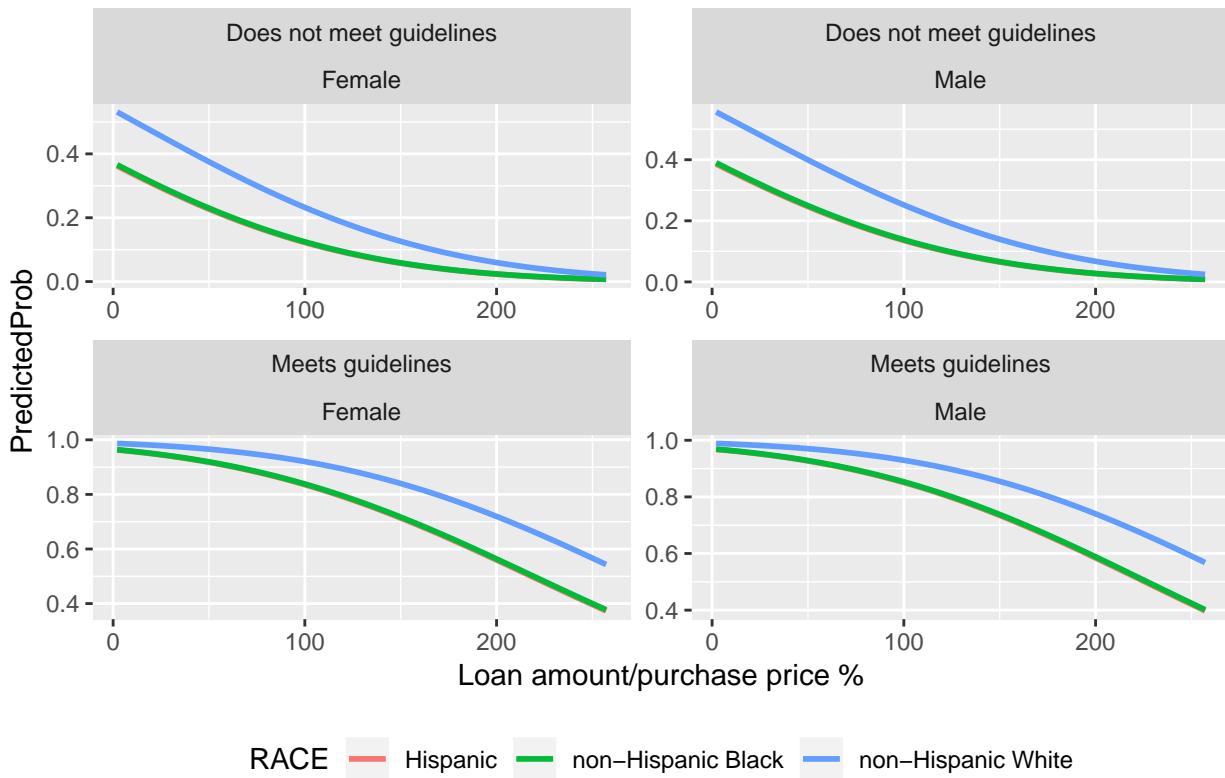
non-Hispanic White Black Predicted probabilities



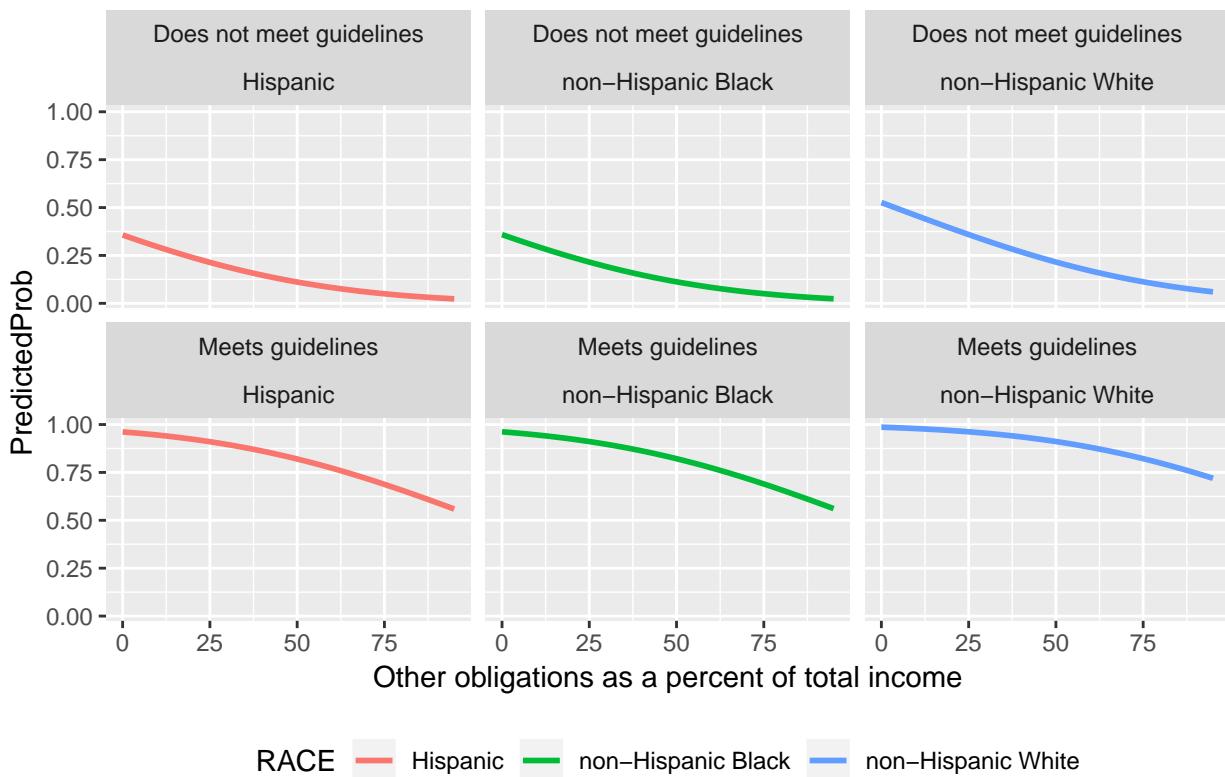
Predicted probabilities



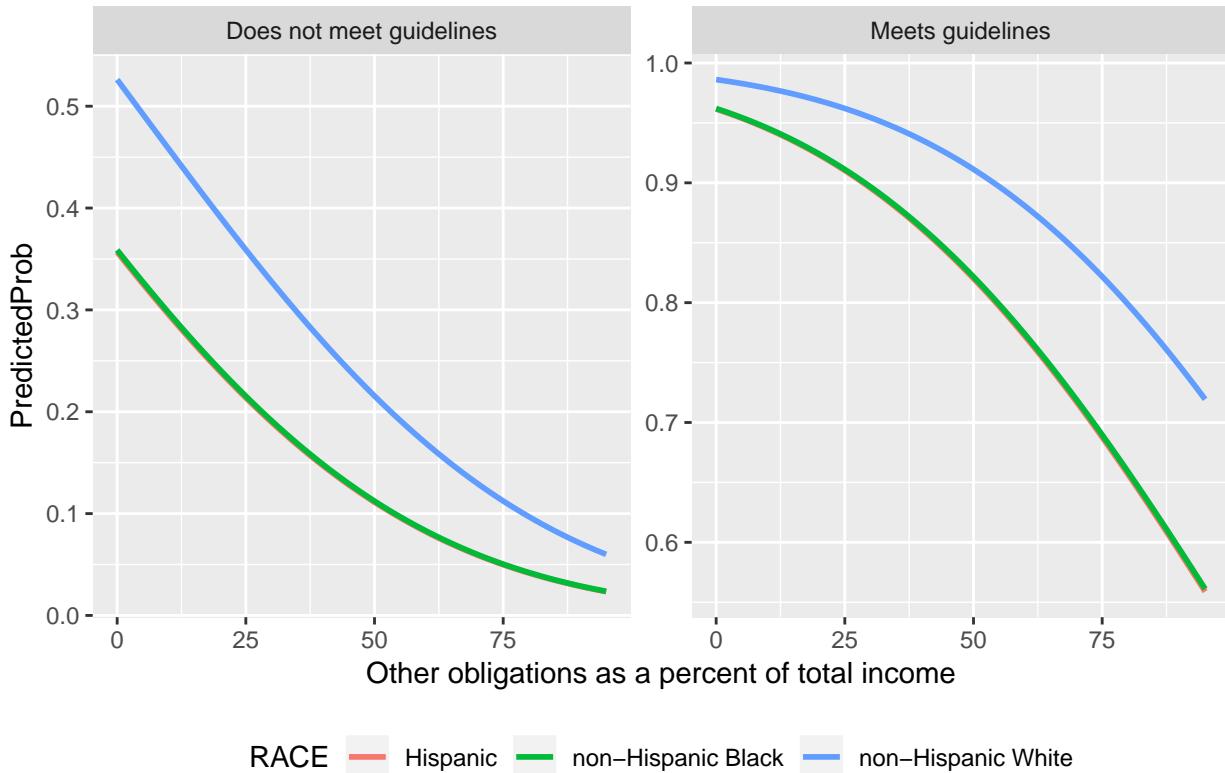
Predicted probabilities



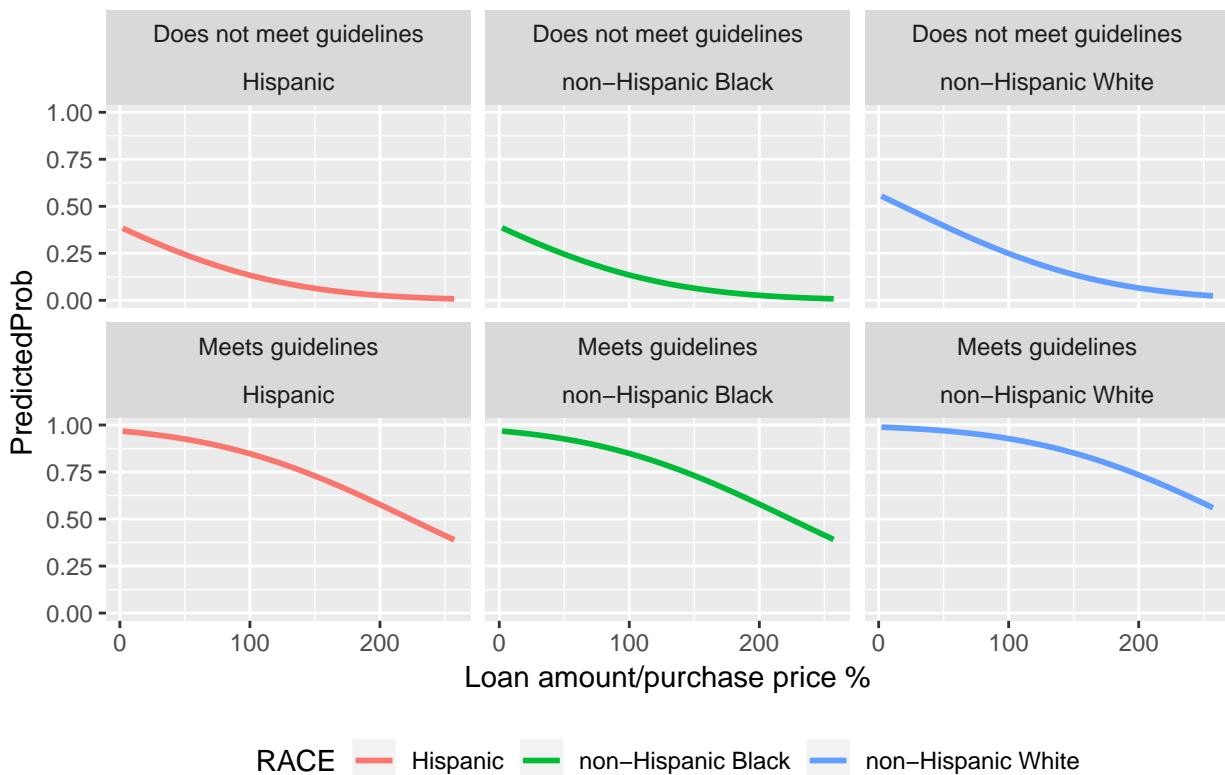
Predicted probabilities



Predicted probabilities



Predicted probabilities



Predicted probabilities

