

EDA Project

Eric Perez

2/21/2022

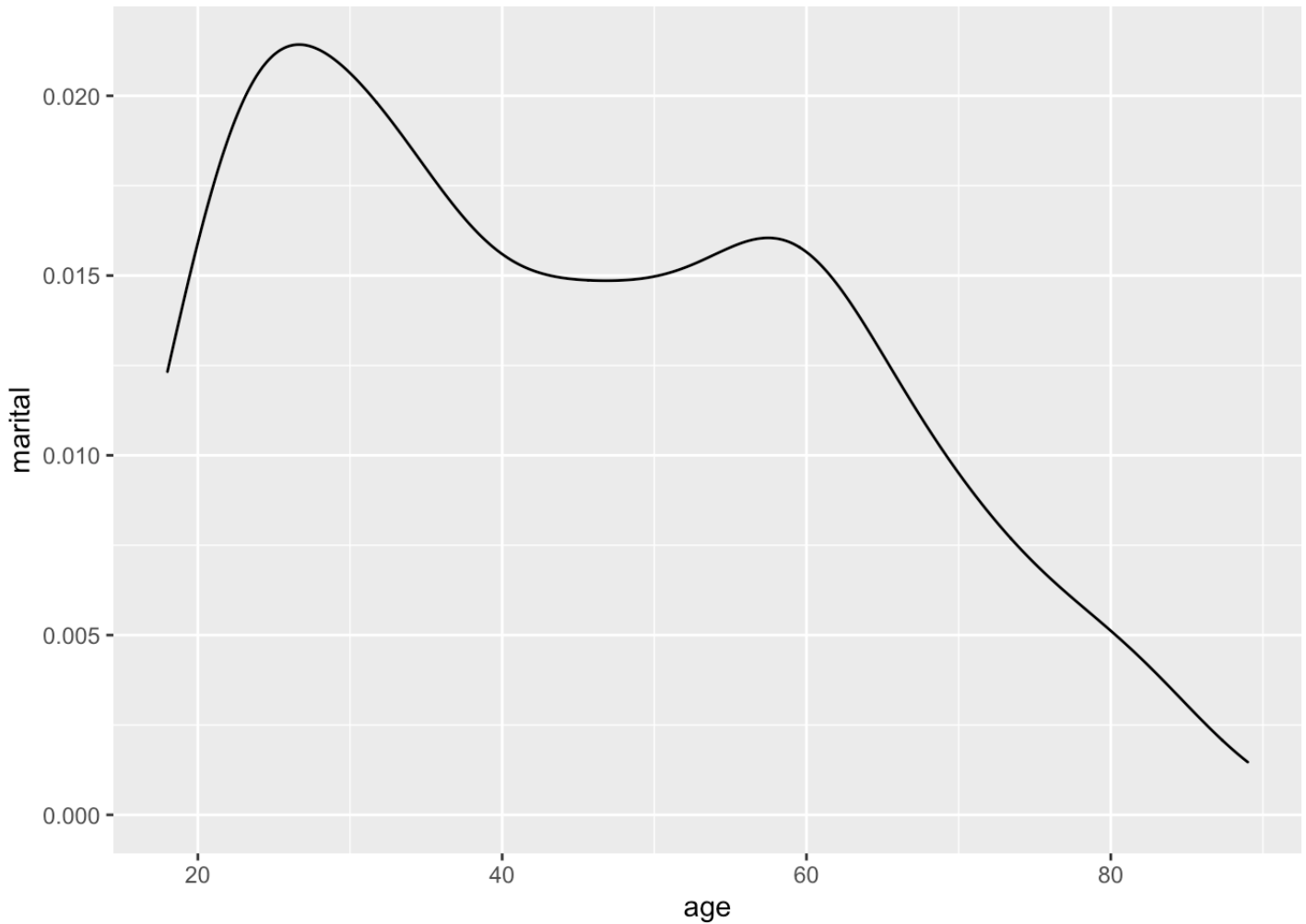
```
depression <-read.delim("/Users/ericperez/Desktop/Homework/depress_081217.txt",header
=TRUE,sep="\t")
dim(depression)
```

```
## [1] 294 37
```

The data basically tells us the ages and their marital status. How many are educated or employed. The data also tells us the income, health, retired, and religion. I want to choose the age and marital status. Those two stood out to me because age plays a big part of marital status. You either stay in a marriage for the sake of it and be depressed or be young and not be married and be content. The true question is what age group is having more marital issues.

Age The ages vary and have all kinds of ages. The lowest age in the data table is 21 and never been married. Marital The older ages around 45 tend to be widowed, divorced, or married. For example, we see someone at 68 being widowed and having more treatment days. The 68 year old has some High School and is retired. The ones with that aren't widowed have less depression because they didn't lose a significant other.

```
ggplot(depression,aes(x=age))+geom_density()+ylab("marital")
```



```
main='Distrubition of age'
```

```
summary(depression$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  18.00   28.00   42.50   44.41   59.00   89.00
```

```
table(depression$age)
```

```
##
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
##  5  5  6  6  9 11  9  6  9  4  5  4  6  5 10  5  9  4  6  5  2  1  5  1  9  7
## 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69
##  2  4  3  6  4  4  5  6  4  2  3  4  3  6  7  9  7  5  4  2  3  5  3  2  4  1
## 70 71 72 73 74 75 77 78 79 80 81 82 83 89
##  5  3  2  2  3  1  2  2  2  1  2  1  5  1
```

```
table(depression$age,useNA="always")
```

```
##
##  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33
##   5   5   6   6   9  11   9   6   9   4   5   4   6   5  10   5
##  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49
##   9   4   6   5   2   1   5   1   9   7   2   4   3   6   4   4
##  50  51  52  53  54  55  56  57  58  59  60  61  62  63  64  65
##   5   6   4   2   3   4   3   6   7   9   7   5   4   2   3   5
##  66  67  68  69  70  71  72  73  74  75  77  78  79  80  81  82
##   3   2   4   1   5   3   2   2   3   1   2   2   2   1   2   1
##  83  89 <NA>
##   5   1   0
```

```
mean(depression$age)
```

```
## [1] 44.41497
```

```
table(depression$age,depression$marital)
```

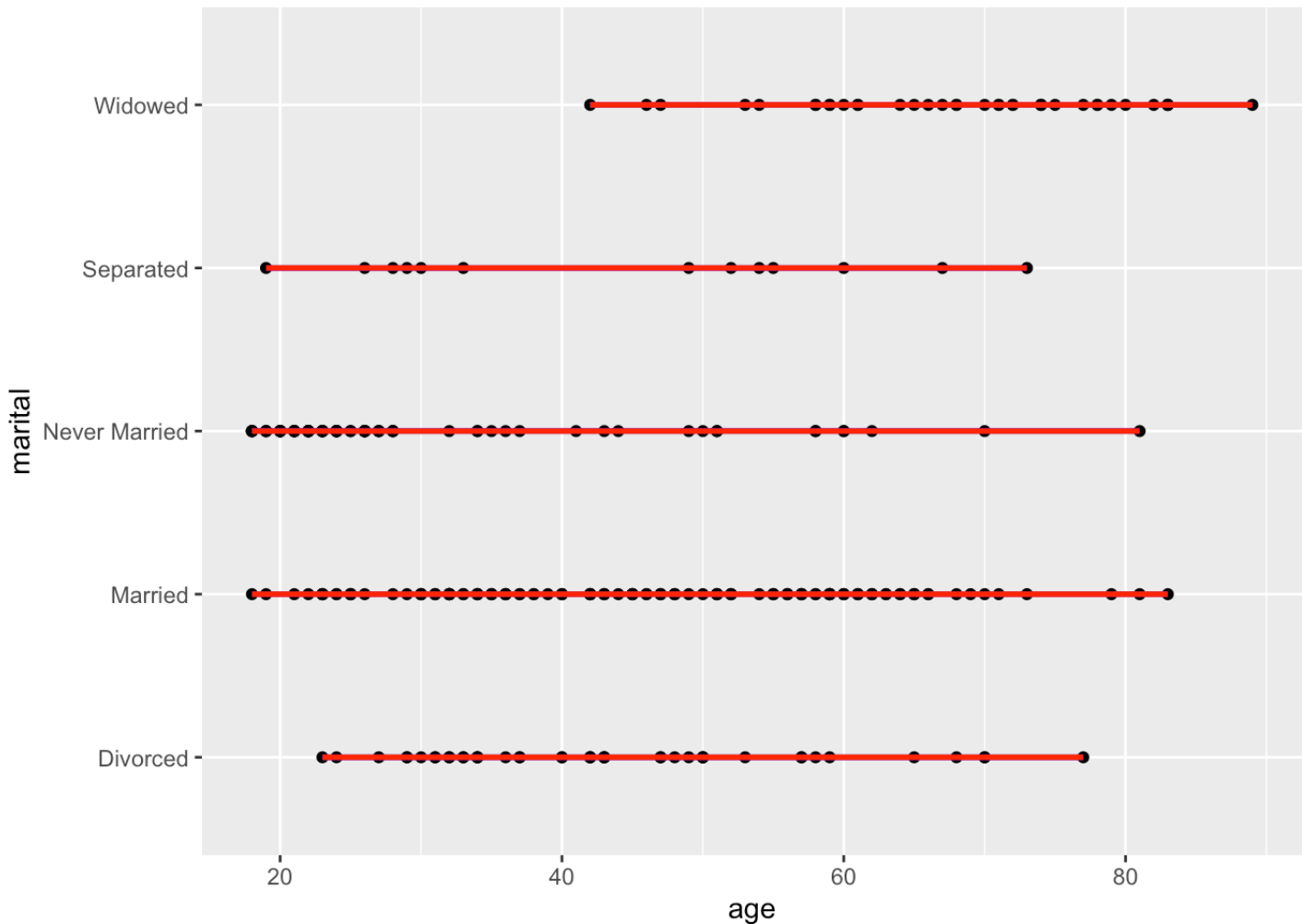
```
##
##      Divorced Married Never Married Separated Widowed
##  18           0       1              4           0       0
##  19           0       1              3           1       0
##  20           0       0              6           0       0
##  21           0       1              5           0       0
##  22           0       1              8           0       0
##  23           1       4              6           0       0
##  24           1       2              6           0       0
##  25           0       4              2           0       0
##  26           0       1              7           1       0
##  27           1       0              3           0       0
##  28           0       1              3           1       0
##  29           2       1              0           1       0
##  30           1       4              0           1       0
##  31           2       3              0           0       0
##  32           2       7              1           0       0
##  33           2       2              0           1       0
##  34           3       4              2           0       0
##  35           0       3              1           0       0
##  36           1       4              1           0       0
##  37           2       2              1           0       0
##  38           0       2              0           0       0
```

##	39	0	1	0	0	0
##	40	2	3	0	0	0
##	41	0	0	1	0	0
##	42	3	5	0	0	1
##	43	3	3	1	0	0
##	44	0	1	1	0	0
##	45	0	4	0	0	0
##	46	0	2	0	0	1
##	47	2	3	0	0	1
##	48	1	3	0	0	0
##	49	1	1	1	1	0
##	50	3	1	1	0	0
##	51	0	4	2	0	0
##	52	0	3	0	1	0
##	53	1	0	0	0	1
##	54	0	1	0	1	1
##	55	0	3	0	1	0
##	56	0	3	0	0	0
##	57	2	4	0	0	0
##	58	1	2	2	0	2
##	59	1	6	0	0	2
##	60	0	3	2	1	1
##	61	0	3	0	0	2
##	62	0	3	1	0	0
##	63	0	2	0	0	0
##	64	0	2	0	0	1
##	65	1	3	0	0	1
##	66	0	2	0	0	1
##	67	0	0	0	1	1
##	68	1	1	0	0	2
##	69	0	1	0	0	0
##	70	2	1	1	0	1
##	71	0	1	0	0	2
##	72	0	0	0	0	2
##	73	0	1	0	1	0
##	74	0	0	0	0	3
##	75	0	0	0	0	1
##	77	1	0	0	0	1
##	78	0	0	0	0	2
##	79	0	1	0	0	1
##	80	0	0	0	0	1
##	81	0	1	1	0	0
##	82	0	0	0	0	1
##	83	0	1	0	0	4
##	89	0	0	0	0	1

```
ggplot(depression,aes(x=age,y=marital))+geom_point()+geom_smooth(se=FALSE,method="lm",
,color="blue")+geom_smooth(se=FALSE,color="RED")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



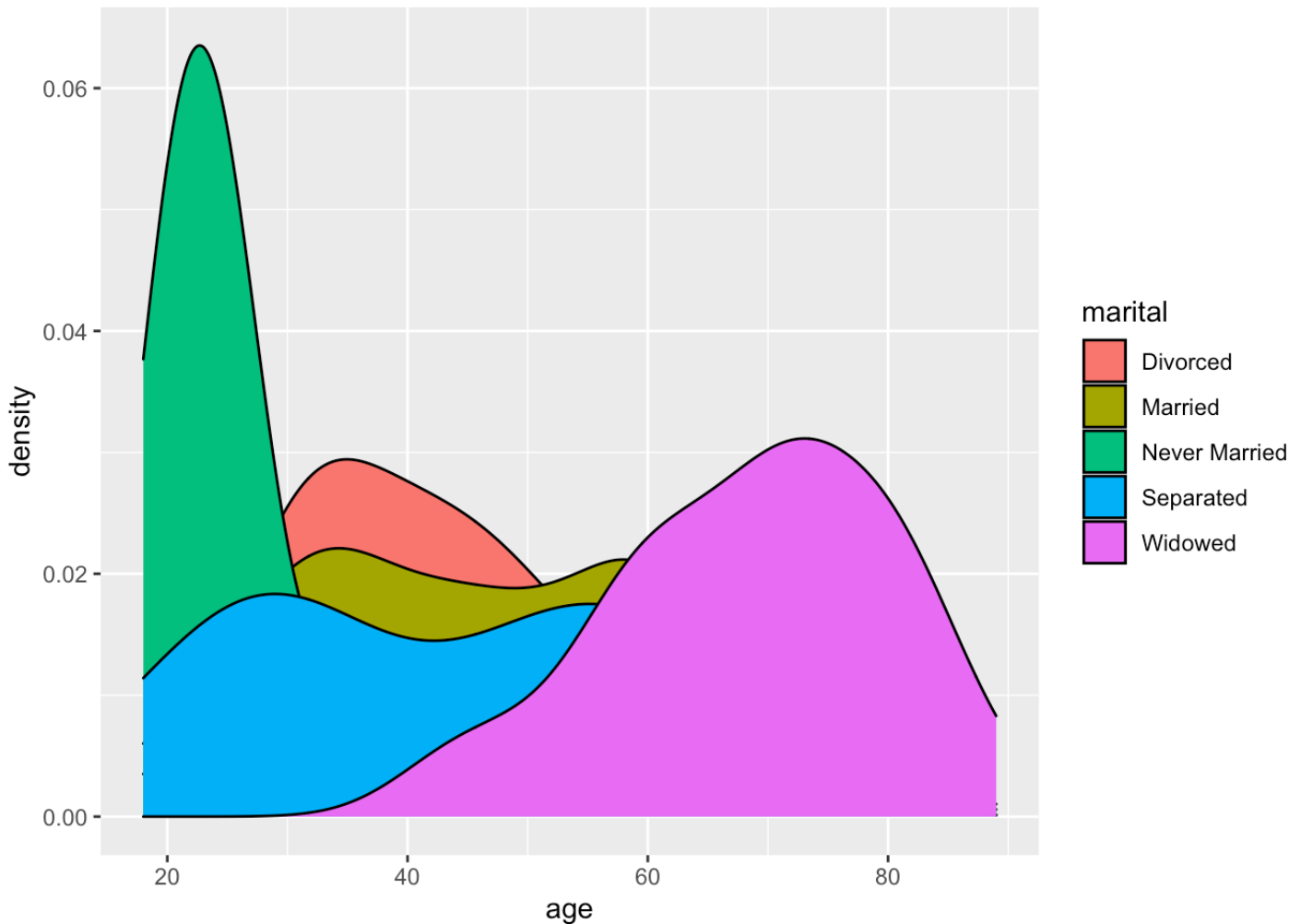
```
table(depression$age,depression$marital)%>% prop.table(margin=2)%>%round(3)
```

```
##
##      Divorced Married Never Married Separated Widowed
##  18      0.000   0.008           0.055    0.000   0.000
##  19      0.000   0.008           0.041    0.077   0.000
##  20      0.000   0.000           0.082    0.000   0.000
##  21      0.000   0.008           0.068    0.000   0.000
##  22      0.000   0.008           0.110    0.000   0.000
```

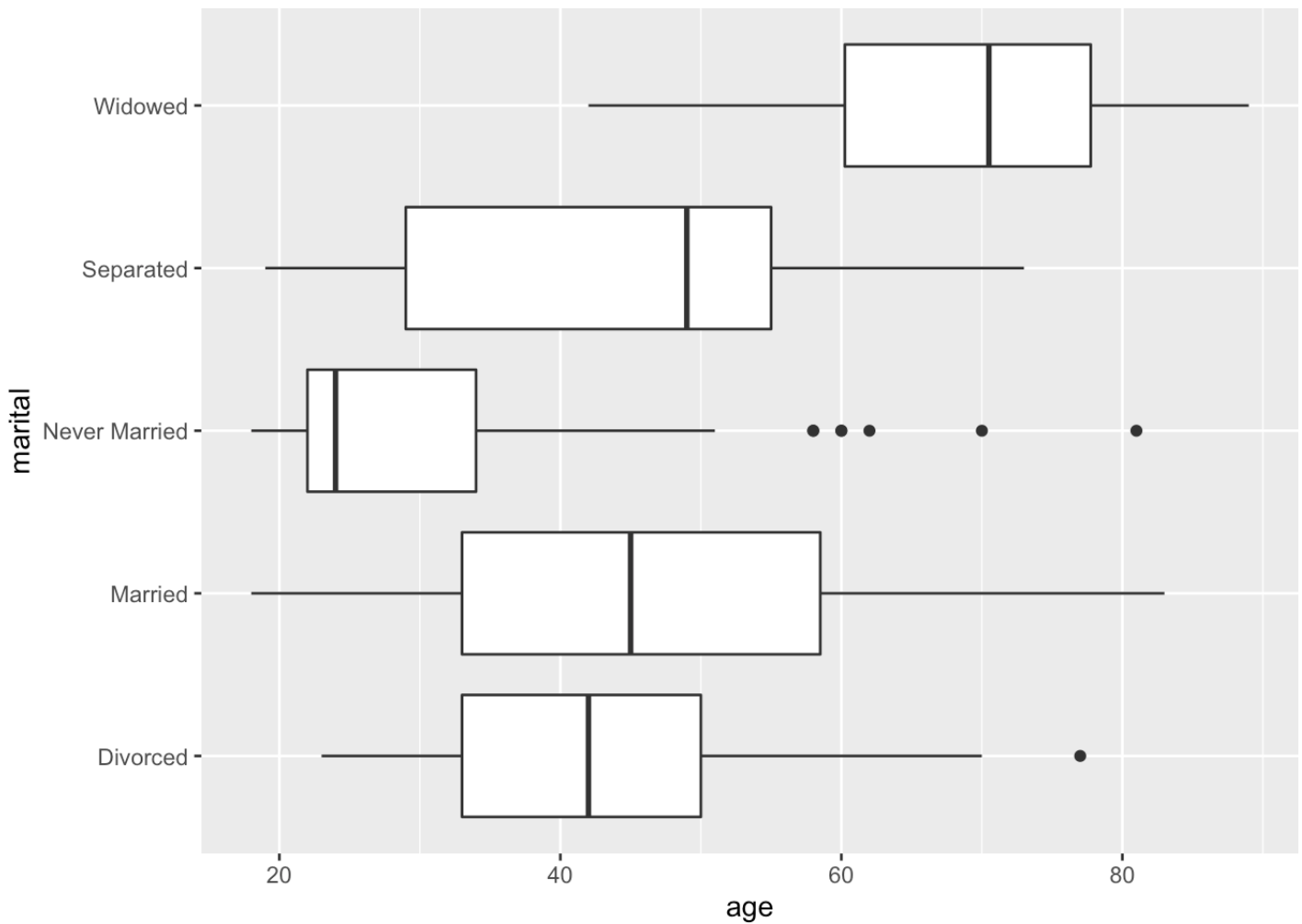
##	23	0.023	0.031	0.082	0.000	0.000
##	24	0.023	0.016	0.082	0.000	0.000
##	25	0.000	0.031	0.027	0.000	0.000
##	26	0.000	0.008	0.096	0.077	0.000
##	27	0.023	0.000	0.041	0.000	0.000
##	28	0.000	0.008	0.041	0.077	0.000
##	29	0.047	0.008	0.000	0.077	0.000
##	30	0.023	0.031	0.000	0.077	0.000
##	31	0.047	0.024	0.000	0.000	0.000
##	32	0.047	0.055	0.014	0.000	0.000
##	33	0.047	0.016	0.000	0.077	0.000
##	34	0.070	0.031	0.027	0.000	0.000
##	35	0.000	0.024	0.014	0.000	0.000
##	36	0.023	0.031	0.014	0.000	0.000
##	37	0.047	0.016	0.014	0.000	0.000
##	38	0.000	0.016	0.000	0.000	0.000
##	39	0.000	0.008	0.000	0.000	0.000
##	40	0.047	0.024	0.000	0.000	0.000
##	41	0.000	0.000	0.014	0.000	0.000
##	42	0.070	0.039	0.000	0.000	0.026
##	43	0.070	0.024	0.014	0.000	0.000
##	44	0.000	0.008	0.014	0.000	0.000
##	45	0.000	0.031	0.000	0.000	0.000
##	46	0.000	0.016	0.000	0.000	0.026
##	47	0.047	0.024	0.000	0.000	0.026
##	48	0.023	0.024	0.000	0.000	0.000
##	49	0.023	0.008	0.014	0.077	0.000
##	50	0.070	0.008	0.014	0.000	0.000
##	51	0.000	0.031	0.027	0.000	0.000
##	52	0.000	0.024	0.000	0.077	0.000
##	53	0.023	0.000	0.000	0.000	0.026
##	54	0.000	0.008	0.000	0.077	0.026
##	55	0.000	0.024	0.000	0.077	0.000
##	56	0.000	0.024	0.000	0.000	0.000
##	57	0.047	0.031	0.000	0.000	0.000
##	58	0.023	0.016	0.027	0.000	0.053
##	59	0.023	0.047	0.000	0.000	0.053
##	60	0.000	0.024	0.027	0.077	0.026
##	61	0.000	0.024	0.000	0.000	0.053
##	62	0.000	0.024	0.014	0.000	0.000
##	63	0.000	0.016	0.000	0.000	0.000
##	64	0.000	0.016	0.000	0.000	0.026
##	65	0.023	0.024	0.000	0.000	0.026
##	66	0.000	0.016	0.000	0.000	0.026
##	67	0.000	0.000	0.000	0.077	0.026
##	68	0.023	0.008	0.000	0.000	0.053
##	69	0.000	0.008	0.000	0.000	0.000

##	70	0.047	0.008	0.014	0.000	0.026
##	71	0.000	0.008	0.000	0.000	0.053
##	72	0.000	0.000	0.000	0.000	0.053
##	73	0.000	0.008	0.000	0.077	0.000
##	74	0.000	0.000	0.000	0.000	0.079
##	75	0.000	0.000	0.000	0.000	0.026
##	77	0.023	0.000	0.000	0.000	0.026
##	78	0.000	0.000	0.000	0.000	0.053
##	79	0.000	0.008	0.000	0.000	0.026
##	80	0.000	0.000	0.000	0.000	0.026
##	81	0.000	0.008	0.014	0.000	0.000
##	82	0.000	0.000	0.000	0.000	0.026
##	83	0.000	0.008	0.000	0.000	0.105
##	89	0.000	0.000	0.000	0.000	0.026

```
ggplot(depression,aes(x=age,fill=marital))+geom_density()
```



```
ggplot(depression,aes(x=age,y=marital))+geom_boxplot()
```



Conclusion I found out that no one between the ages 20-40 years old are widowed. The ages 20-80 are married. I found the most people are divorced between ages of 30 to 40 years old. The people around the ages of late 20's early 30's are separated. There's a couple around the age of 50 that are also separated.