Health Analytics: Exploring the BRFSS Dataset

Author: Bibobra Alabrah

Date: 7/24/2018

In [1]:
```
# Setup
# Load packages
```

In [2]:
```
library(ggplot2)
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

In [3]:
```
# Load data
load("brfss2013.RData")
```

About The Data

This dataset was provided by BRFSS whose objective is to collect uniform, state-specific data acros the 50 states and participating US territories on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases that affect the adult population. Factors assessed by the BRFSS in 2013 include tobacco use, HIV/AIDS knowledge and prevention, exercise, immunization, health status, healthy days - health-related quality of life, health care access, inadequate sleep, hypertension awareness, cholesterol awareness, chronic health conditions, alcohol consumption, fruits and vegetables consumption, arthritis burden, and seatbelt use. The main survey covered 17 compulsory sections viz record identification, health status, healthy days, health care access, inadequate sleep, hypertension awareness, cholesterol awareness, chronic health conditions, demographics, tobacco use, alcohol consumption, fruits and vegetables, exercise, Arthritis burden, use seatbelts, immunization, HIV/AIDS, and 22 optional modules. This dataset has 491,775 observations with 330 variables.

Given that this dataset was collected in all the 50 states and US territories randomly, the results can be generalized to the entire population, however, causality cannot be infered.

Research questions

This is a very interesting and robust dataset with several possible research questions. For curiosity sake, I have chosen these two for consideration.

Research quesion 1: Is a person's chance of getting a heart attack related to regular medical check ups, and lack of money for medicine due to cost? I think is very interesting because routine medical checkups may or may not be that influencial given that these this chronic condition come unexpectedly. Money is always an issue so it will be interesting to see its impact.

Research quesion 2: Is depression disorder prevalent amongst the less educated, low income earners, employment status, married or single, which gender suffers most? This fascinates me because depression cuts across all classes of people but I am sure that there are some correlations.

Exploratory data analysis

In [4]:
```
# Check the data dimension
dim(brfss2013)
```

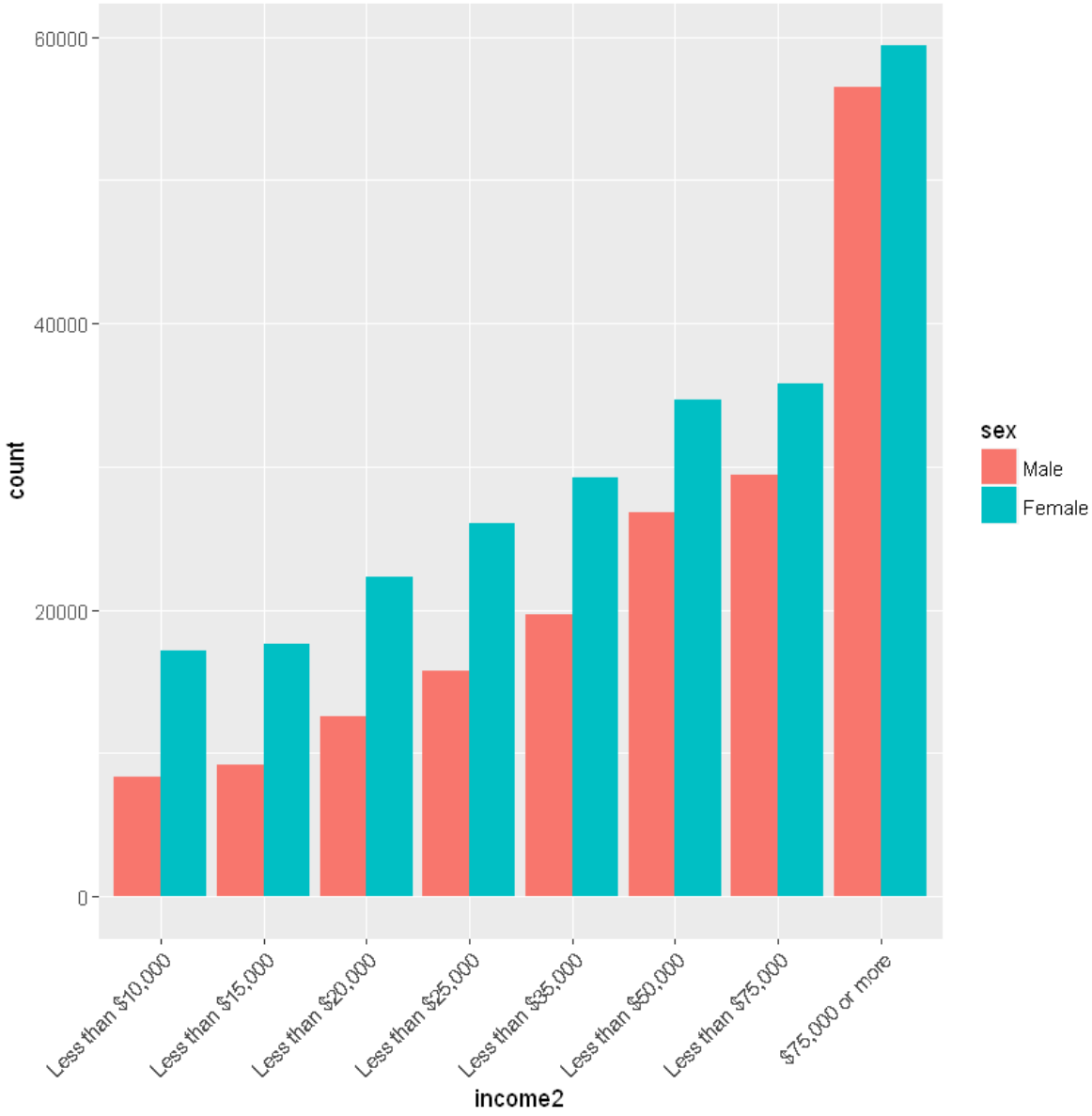
```
491775  330
```

In [5]:
```
# View the first 3 rows
head(brfss2013, 3)
```

X_state	fmonth	idate	imonth	iday	iyear	dispcode	seqno	X_psu	ctel
Alabama	January	1092013	January	9	2013	Completed interview	2013000580	2013000580	Yes
Alabama	January	1192013	January	19	2013	Completed interview	2013000593	2013000593	Yes
Alabama	January	1192013	January	19	2013	Completed interview	2013000600	2013000600	Yes

The data dimension tells us that there are 330 variables. The meanings of these abbreviations are found in the code book found in my github profile.

In [89]:
```
# What is the income level distribution?
brfss2013 %>%
filter(!is.na(income2), !is.na(sex)) %>%
ggplot(aes(x = income2, fill = sex)) +
geom_bar(stat = "count", position = position_dodge()) +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



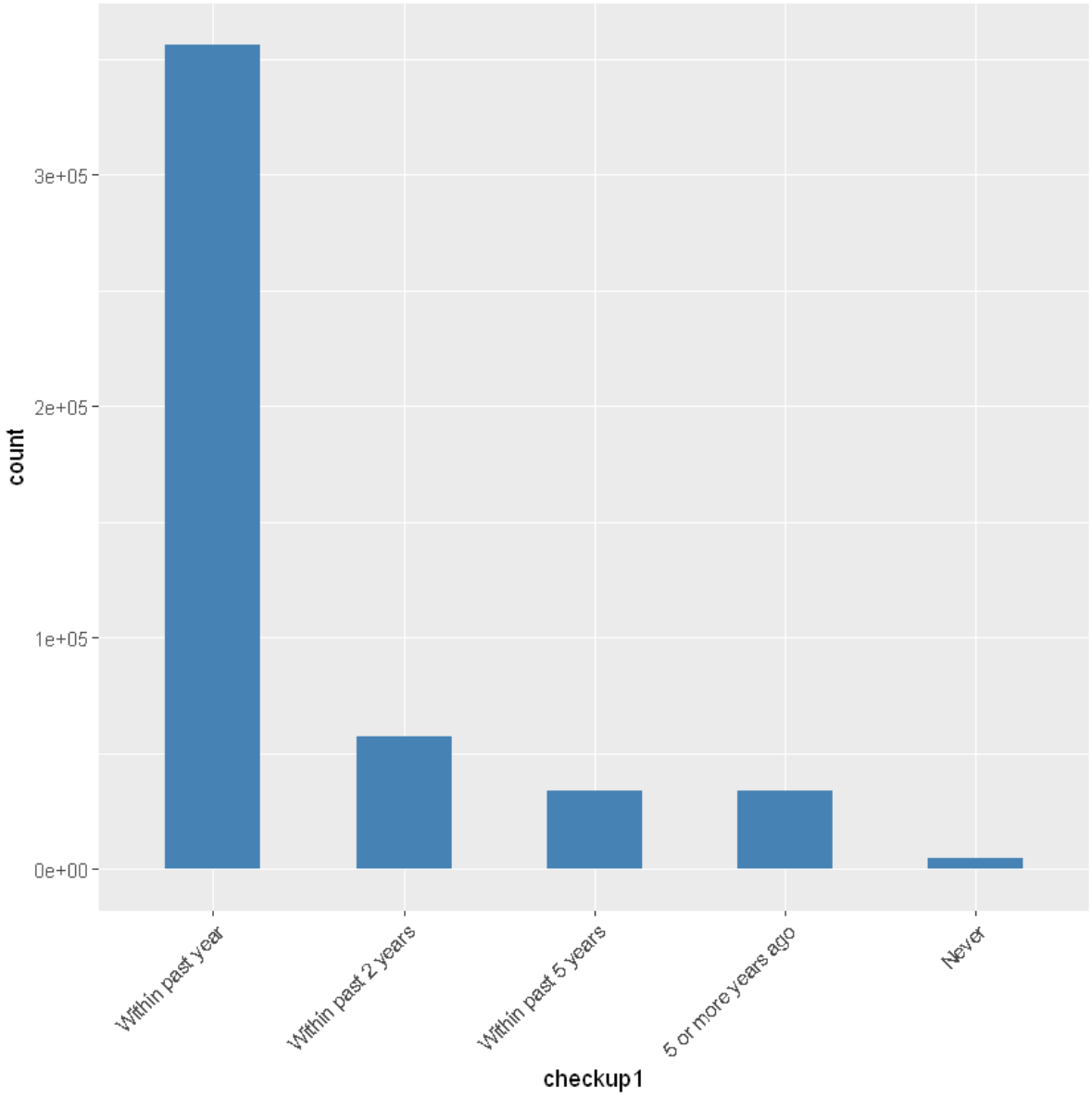
Females made more money than males as can be seen in the plot above. As expected, income increased with education level.

In [7]:
```
# Display the summary stats
round(prop.table(table(brfss2013$income2, brfss2013$sex)), 2)
```

```
                   Male Female
Less than $10,000 0.02   0.04
Less than $15,000 0.02   0.04
Less than $20,000 0.03   0.05
Less than $25,000 0.04   0.06
Less than $35,000 0.05   0.07
Less than $50,000 0.06   0.08
Less than $75,000 0.07   0.09
$75,000 or more   0.13   0.14
```

Pretty much all class of income earners were included in this survey; majority of the respondents earn more than $75,000. From this summary statistic, we could infer that they were more high income level respondents than low income level respondents.

In [80]:
```
# Let us explore the checkup variable
brfss2013 %>%
filter(!is.na(checkup1)) %>%
ggplot(aes(x = checkup1)) +
geom_bar(stat = "count", width = 0.5, fill = "steelblue") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

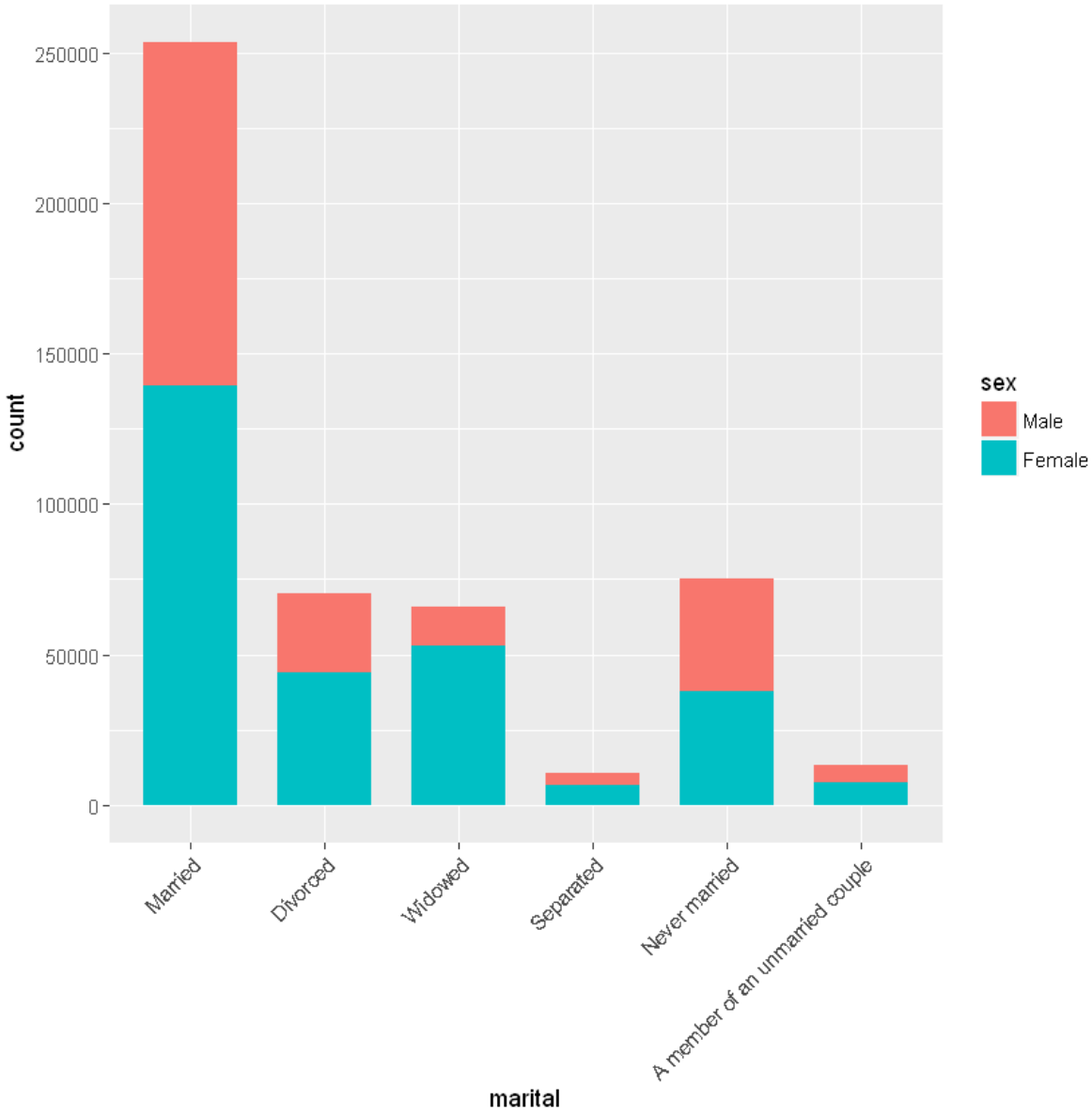


In [9]:
```
# Let us see the numbers
round(prop.table(table(brfss2013$checkup1)), 2)
```

```
   Within past year Within past 2 years Within past 5 years 5 or more years a
go
              0.73               0.12               0.07                0.
07
             Never
              0.01
```

73% did regular check ups within the past one year.

In [84]:
```
# Let us take a look at the marital status variable
brfss2013 %>%
filter(!is.na(marital), !is.na(sex)) %>%
ggplot(aes(x = marital, fill = sex)) +
geom_bar(width = 0.7, stat = "count") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



In [11]:
```
# Let us pull up the summary statistics
round(prop.table(table(brfss2013$marital)), 2)
```

```
           Married           Divorced
              0.52               0.14
           Widowed          Separated
              0.13               0.02
    Never married A member of an unmarried couple
              0.15               0.03
```

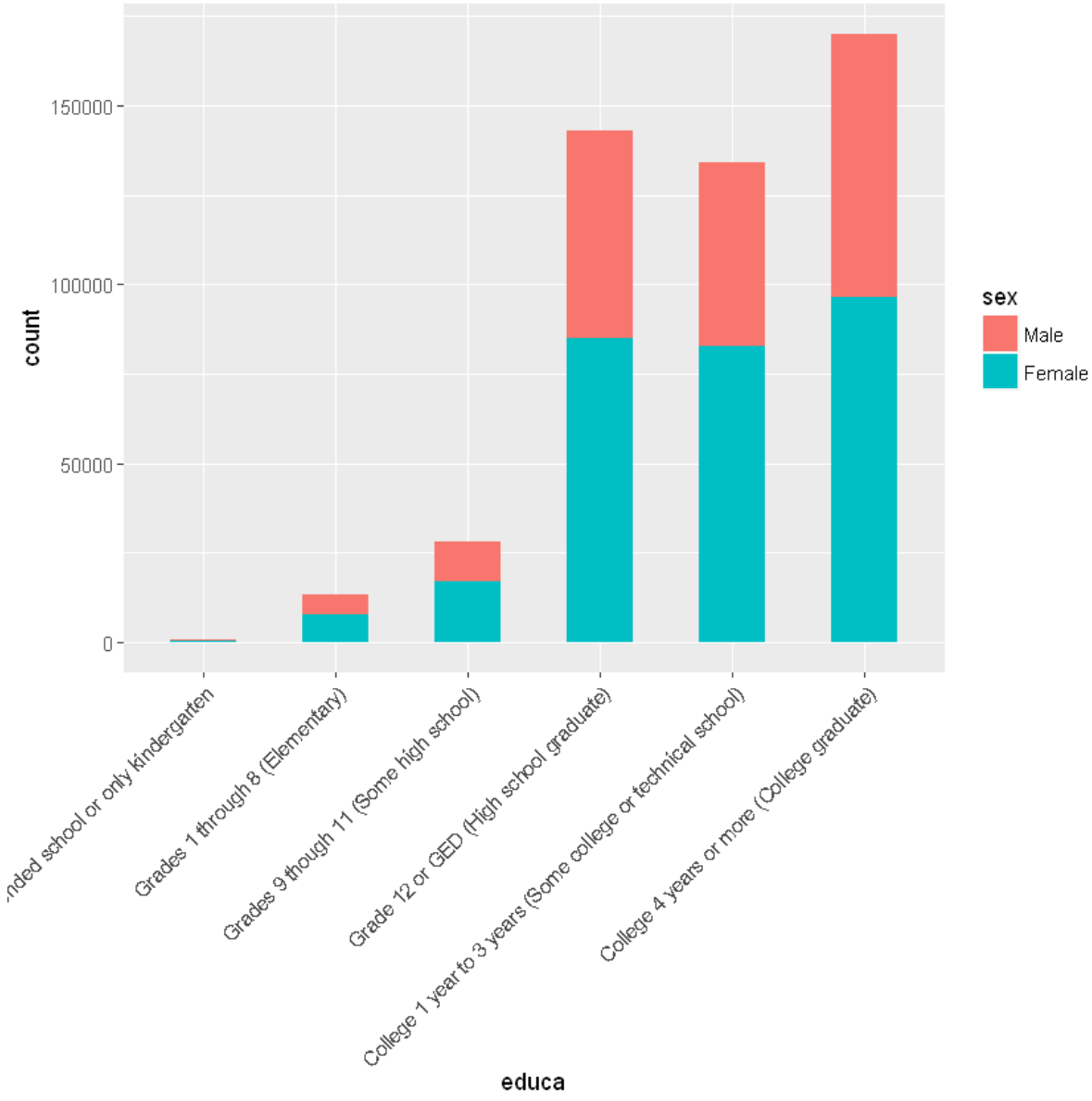
There are more married people, followed by divorced, and then never married.

In [12]:
```
# let us explore the education level of the respondents
round(prop.table(table(brfss2013$educa)), 2)
```

```
                                                  Never attended school or only kindergarten
                                                                                        0.00
                                                            Grades 1 through 8 (Elementary)
                                                                                        0.03
                                                       Grades 9 though 11 (Some high school)
                                                                                        0.06
                                                  Grade 12 or GED (High school graduate)
                                                                                        0.29
College 1 year to 3 years (Some college or technical school)
                                                                                        0.27
                                                  College 4 years or more (College graduate)
                                                                                        0.35
```

37% of the respondents were college 4 years or more graduates, this is followed by High school grads with 29%, and then technical school grads with 27%.

In [81]:
```
brfss2013 %>%
filter(!is.na(educa), !is.na(sex)) %>%
ggplot(aes(x = educa, fill = sex)) +
geom_bar(stat = "count", width = 0.5) +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

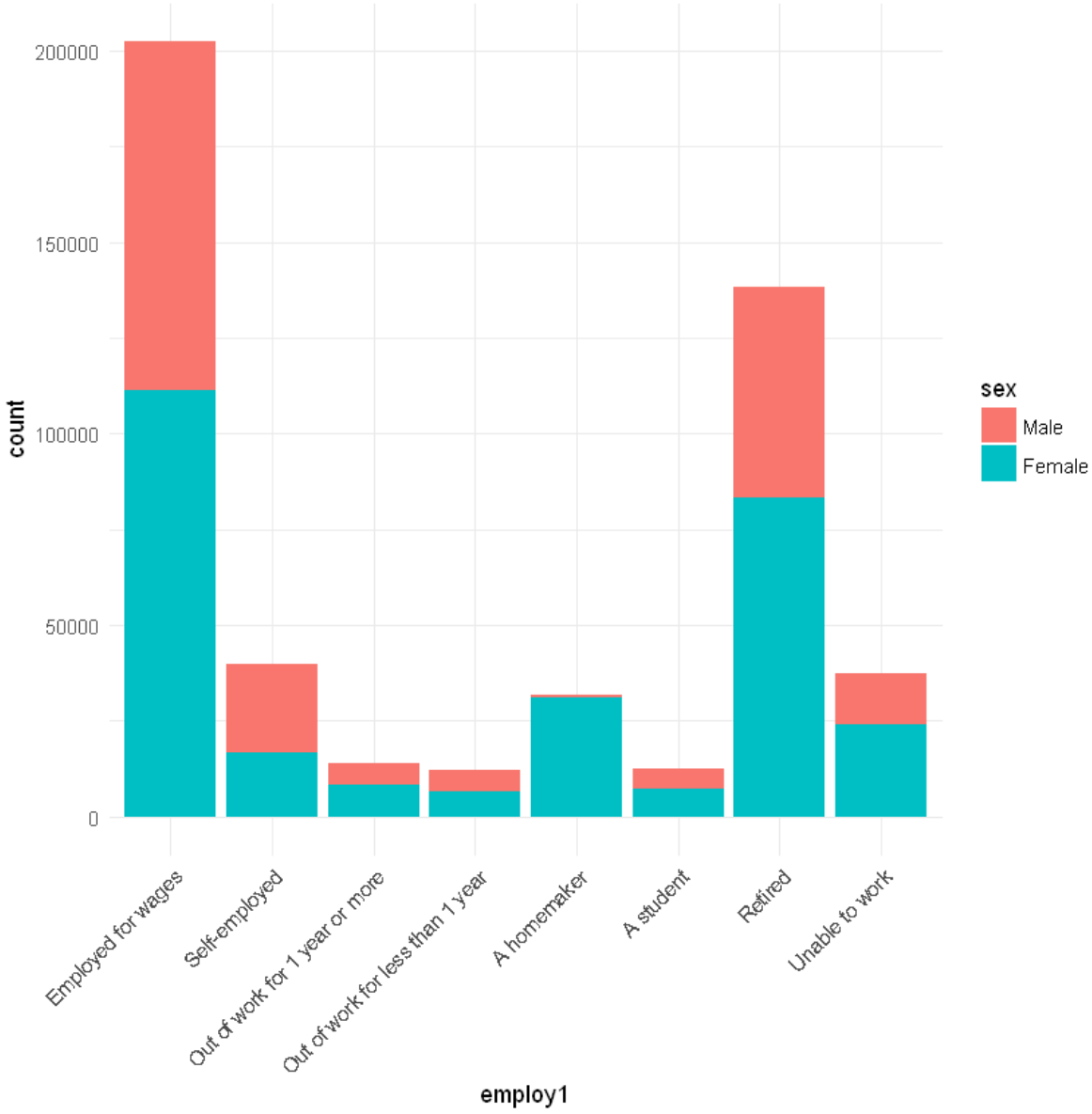


More females had education relative to males.

In [14]:
```
# The employment variable might play a in depression disorder. Let us explore
 this feature
round(prop.table(table(brfss2013$employ1, brfss2013$sex)), 2)
```

	Male	Female
Employed for wages	0.19	0.23
Self-employed	0.05	0.03
Out of work for 1 year or more	0.01	0.02
Out of work for less than 1 year	0.01	0.01
A homemaker	0.00	0.06
A student	0.01	0.01
Retired	0.11	0.17
Unable to work	0.03	0.05

In [87]:
```
brfss2013 %>%
filter(!is.na(employ1), !is.na(sex)) %>%
ggplot(aes(x = employ1, fill = sex)) +
geom_bar(stat = "count") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



More females are employed compared to men, however, there were more men in the self-employed category. This is not surprising because earlier we see that females made more money across board.

Research quesion 1:

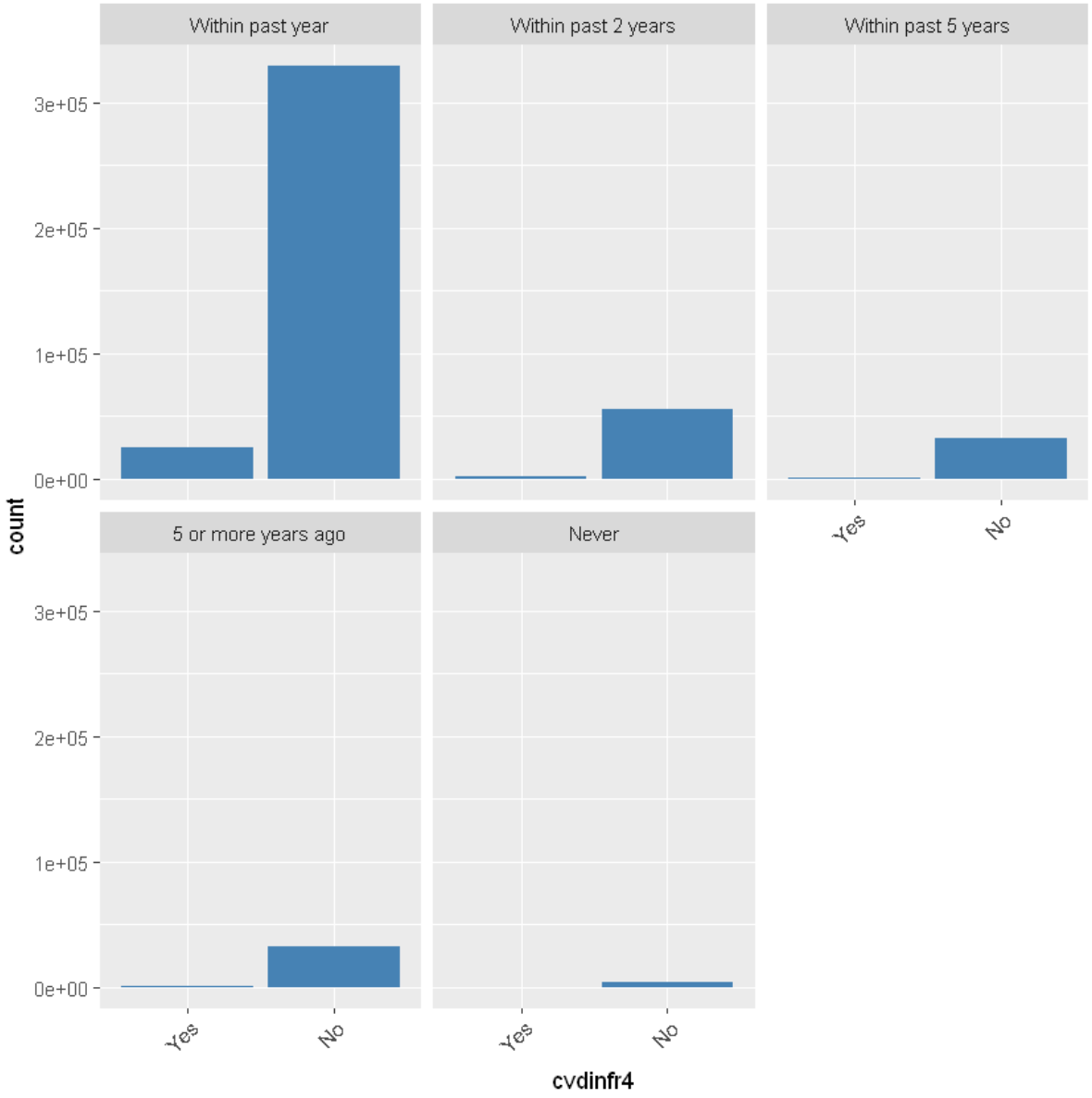
Is a person's chance of getting a heart attack related to regular medical check ups, and lack of money for medicine due to cost? I think is very interesting because routine medical checkups may or may not be that influencial given that these this chronic condition come unexpectedly. Money is always an issue so it will be interesting to see its impact.

I will explore the heart attack variable against the check up and no money due to medicine cost variable and investigate any possible correlations.

In [90]:
```
brfss2013 %>%
    filter(!is.na(cvdinfr4), !is.na(checkup1)) %>%
    group_by(cvdinfr4, checkup1) %>%
    summarise(count = n())

# cvdinfr4 stands for "Ever Diagnosed With Heart Attack?"
```

cvdinfr4	checkup1	count
Yes	Within past year	24772
Yes	Within past 2 years	1869
Yes	Within past 5 years	943
Yes	5 or more years ago	1058
Yes	Never	199
No	Within past year	329546
No	Within past 2 years	55215
No	Within past 5 years	32596
No	5 or more years ago	32523
No	Never	4291

In [93]:
```
brfss2013 %>%
filter(!is.na(cvdinfr4), !is.na(checkup1)) %>%
ggplot(aes(x = cvdinfr4)) +
geom_bar(fill = "steelblue", stat = "count") +
facet_wrap(~ checkup1) +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

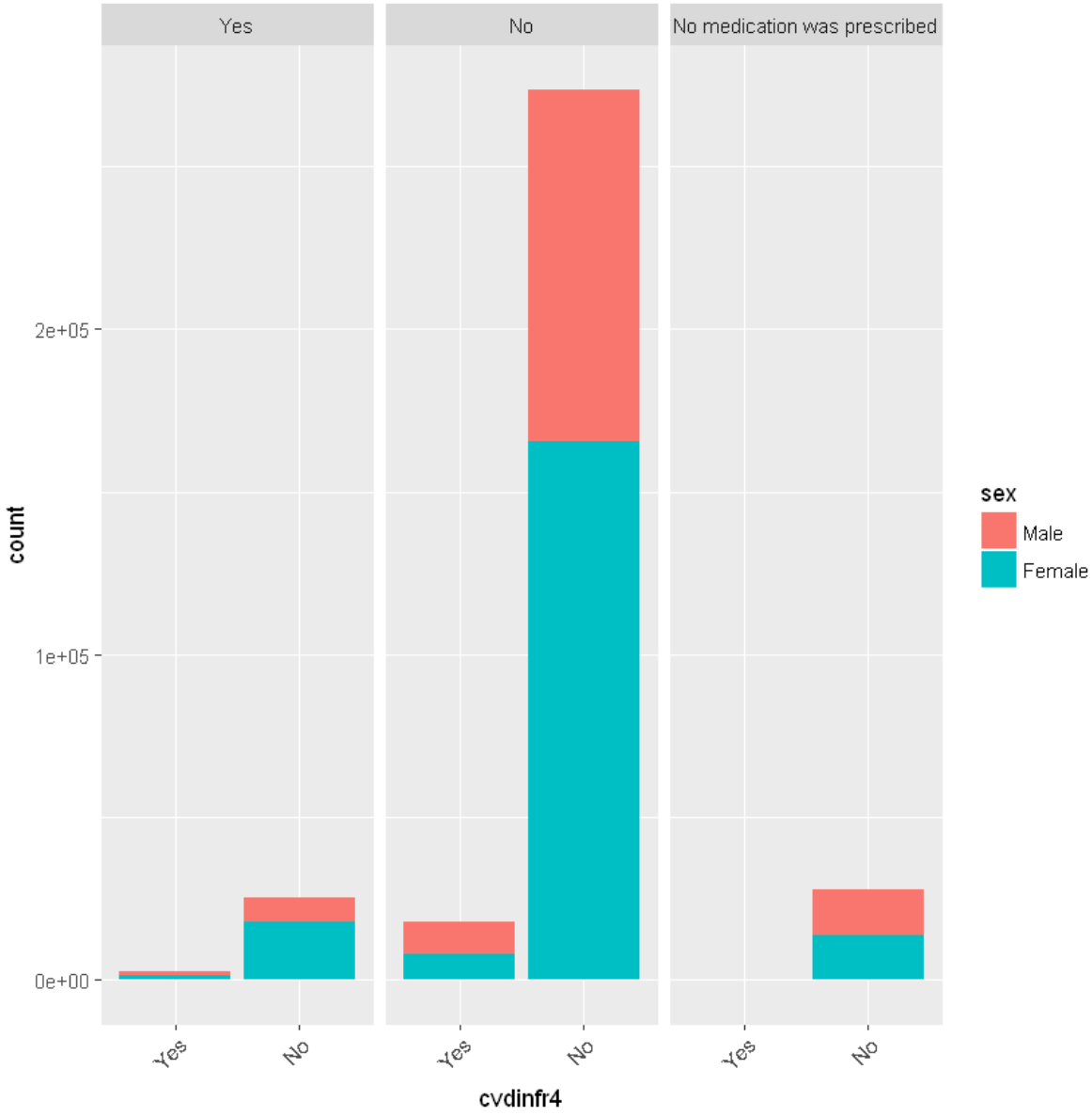


We can see that there is a correlation between the check up and heart attack. Majority of those who went for medical check up within the past year were free from this chronic condition. It makes sense to say that, visiting the doctor regularly is a good way to avoid heart attack.

In [91]:
```
# Let us now look at the no money to pay for meds variable.
brfss2013 %>%
filter(!is.na(cvdinfr4)) %>%
group_by(cvdinfr4, medscost) %>%
summarise(count = n())
```

cvdinfr4	medscost	count
Yes	Yes	2650
Yes	No	17988
Yes	No medication was prescribed	332
Yes	NA	8314
No	Yes	25404
No	No	273376
No	No medication was prescribed	27619
No	NA	133505

In [95]:
```
brfss2013 %>%
filter(!is.na(cvdinfr4), !is.na(sex), !is.na(medscost)) %>%
ggplot(aes(x = cvdinfr4, fill = sex)) +
geom_bar(stat = "count") +
facet_wrap(~ medscost) +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



We can see that for majority of the respondents money to pay for medicine was not the issue but yet they were diagonized of this condition. Hence, there are other causal effects, we can only look at correlations with respect to this matter.

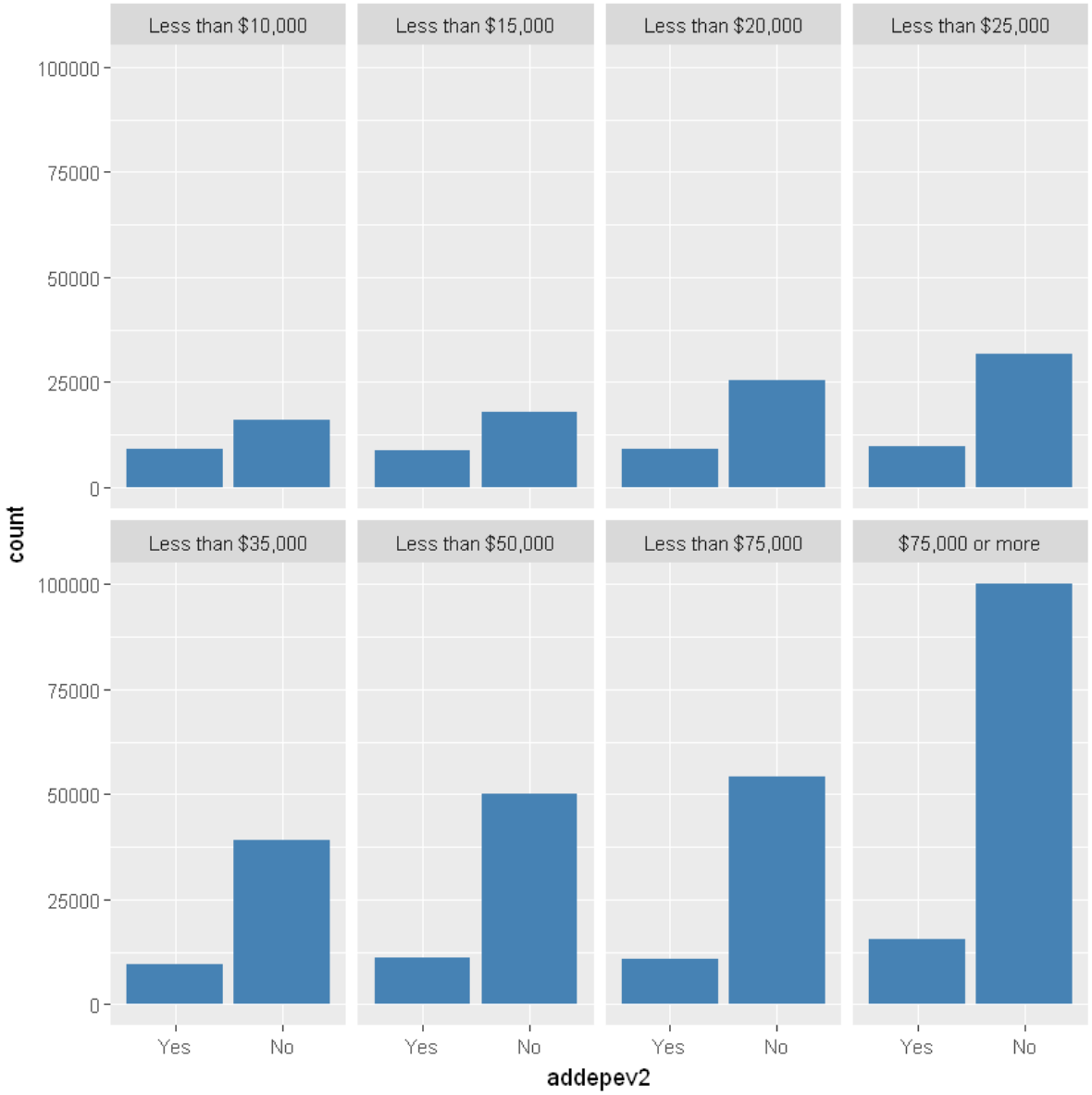
Research quesion 2: Is depression disorder prevalent amongst the less educated, low income earners, employment status, married or single, which gender suffers most? This fascinates me because depression cuts across all classes of people but I am sure that there are some correlations.

In [96]:
```
# effects of education level on those who had depression disorder
brfss2013 %>%
    filter(addepev2 == 'Yes', !is.na(educa)) %>%
    group_by(addepev2, educa) %>%
    summarise(count = n())
```

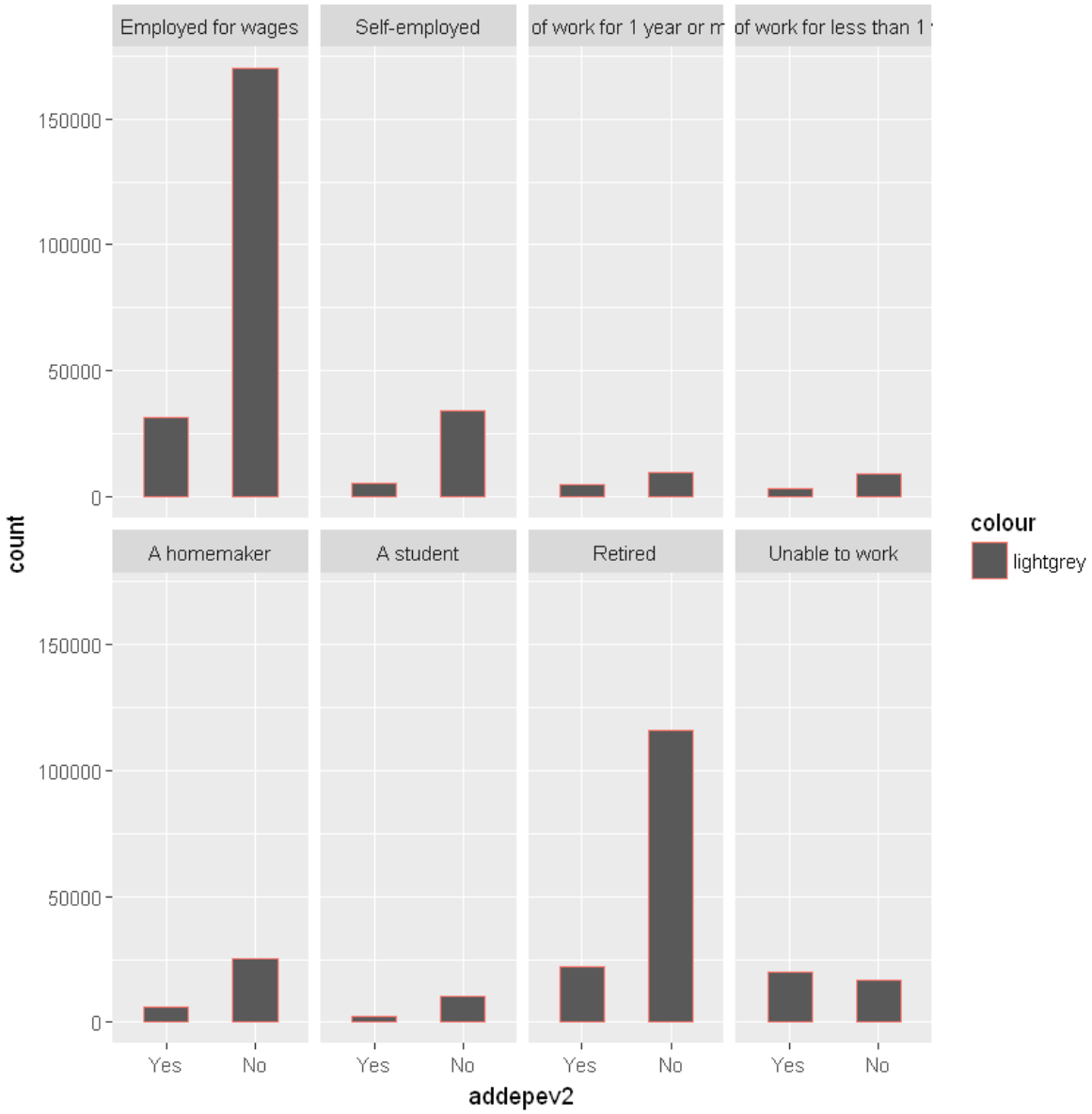
addepev2	educa	count
Yes	Never attended school or only kindergarten	139
Yes	Grades 1 through 8 (Elementary)	3030
Yes	Grades 9 though 11 (Some high school)	7432
Yes	Grade 12 or GED (High school graduate)	27933
Yes	College 1 year to 3 years (Some college or technical school)	29050
Yes	College 4 years or more (College graduate)	27904

We can see that those with higher education suffer more from depression disorder.

In [59]:
```
brfss2013 %>%
filter(!is.na(addepev2), !is.na(income2)) %>%
group_by(income2) %>%
ggplot(aes(x = addepev2)) +
geom_bar(stat = "count", fill ="steelblue") +
facet_wrap(~ income2, ncol = 4)
```



In [110]:
```
# What impact does employment status has on depression disorder?

brfss2013 %>%
filter(!is.na(addepev2), !is.na(employ1), !is.na(sex)) %>%
group_by(employ1) %>%
ggplot(aes(x = addepev2, color = "lightgrey")) +
geom_bar(stat = "count", width = 0.5) +
facet_wrap(~ employ1, ncol = 4)
```

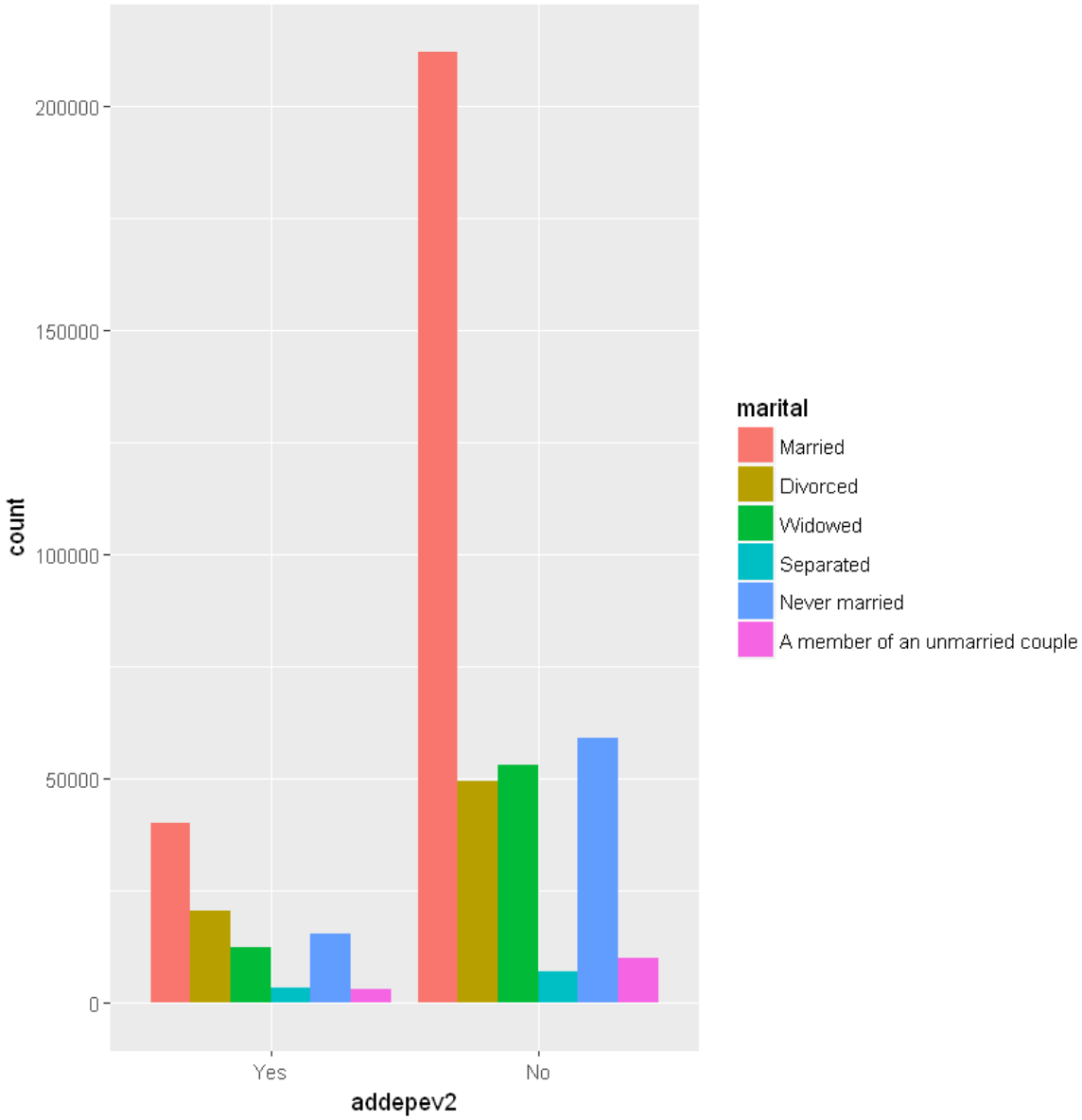


In [113]:
```
brfss2013 %>%
filter(!is.na(cvdinfr4), !is.na(employ1)) %>%
group_by(cvdinfr4, employ1) %>%
summarise(count = n())
```

cvdinfr4	employ1	count
Yes	Employed for wages	4221
Yes	Self-employed	1385
Yes	Out of work for 1 year or more	748
Yes	Out of work for less than 1 year	390
Yes	A homemaker	1292
Yes	A student	79
Yes	Retired	15371
Yes	Unable to work	5626
No	Employed for wages	197473
No	Self-employed	38328
No	Out of work for 1 year or more	13216
No	Out of work for less than 1 year	11795
No	A homemaker	30207
No	A student	12582
No	Retired	121876
No	Unable to work	31283

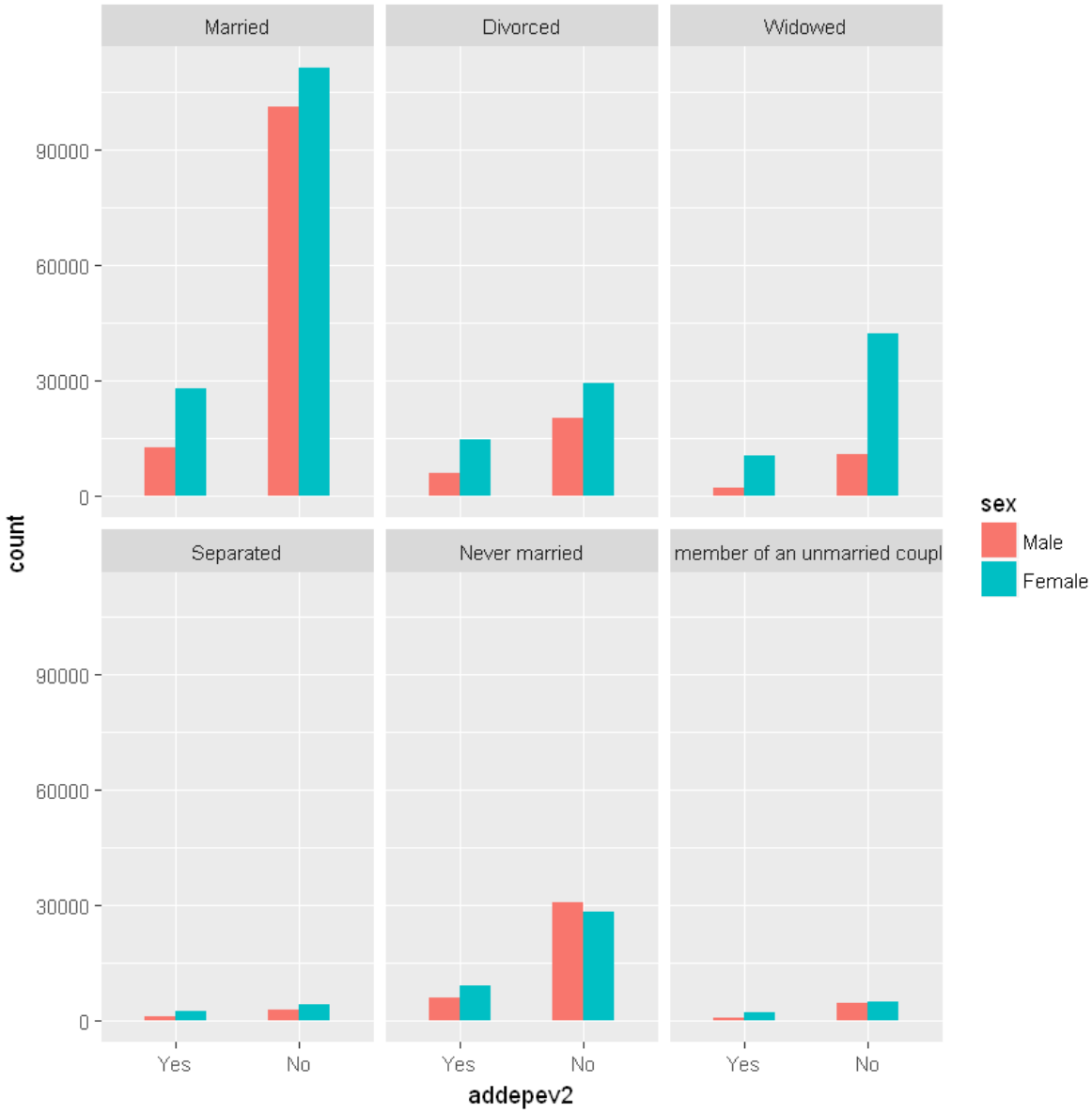
It seems like depression disorder is independent of employment status. Majority of those employed for wages were not depressed, however, some were, maybe they hate their jobs. This also applied to those who are self-employed, maybe those depressed at having a hard time growing there businesses. Majority of the students were not depressed.

In [123]:
```
# Finally, let us explore the relationship between marital status and depressi
on cases.

brfss2013 %>%
filter(!is.na(addepev2), !is.na(marital)) %>%
ggplot(aes(x = addepev2, fill = marital)) +
  geom_bar(stat = "count", position = position_dodge())
```



Those who were married suffer depression disorder more than any other category, this was followed by divorcees. The flip side looks good though.

In [127]:
```
# Which gender suffered more dispression disorder?

brfss2013 %>%
filter(!is.na(addepev2), !is.na(sex), !is.na(marital)) %>%
ggplot(aes(x = addepev2, fill = sex)) +
geom_bar(stat = "count", position = position_dodge(), width = 0.5) +
facet_wrap(~ marital)
```



CONCLUSION

1. Majority of those who went for medical check up within the past year were free from heart attack. It makes sense to say that, visiting the doctor regularly is a good way to avoid heart attack.
2. For majority of the respondents, money to pay for medicine was not an issue but yet they were diagonized of this condition. Hence, there are other causal effects, we can only look at correlations with respect to this matter.
3. Respondents with higher education suffer more from depression disorder.
4. It seems like depression disorder is independent of employment status. Majority of those employed for wages were not depressed, however, some were, maybe they hate their jobs. This also applied to those who are self-employed, maybe those depressed at having a hard time growing there businesses. Majority of the students were not depressed.
5. Those who were married suffer depression disorder more than any other category, this was followed by divorcees. The flip side looks good though. Males suffer less depression disorder.