Assignment4-Bank DF Insights(ggplot)

GundaSuman

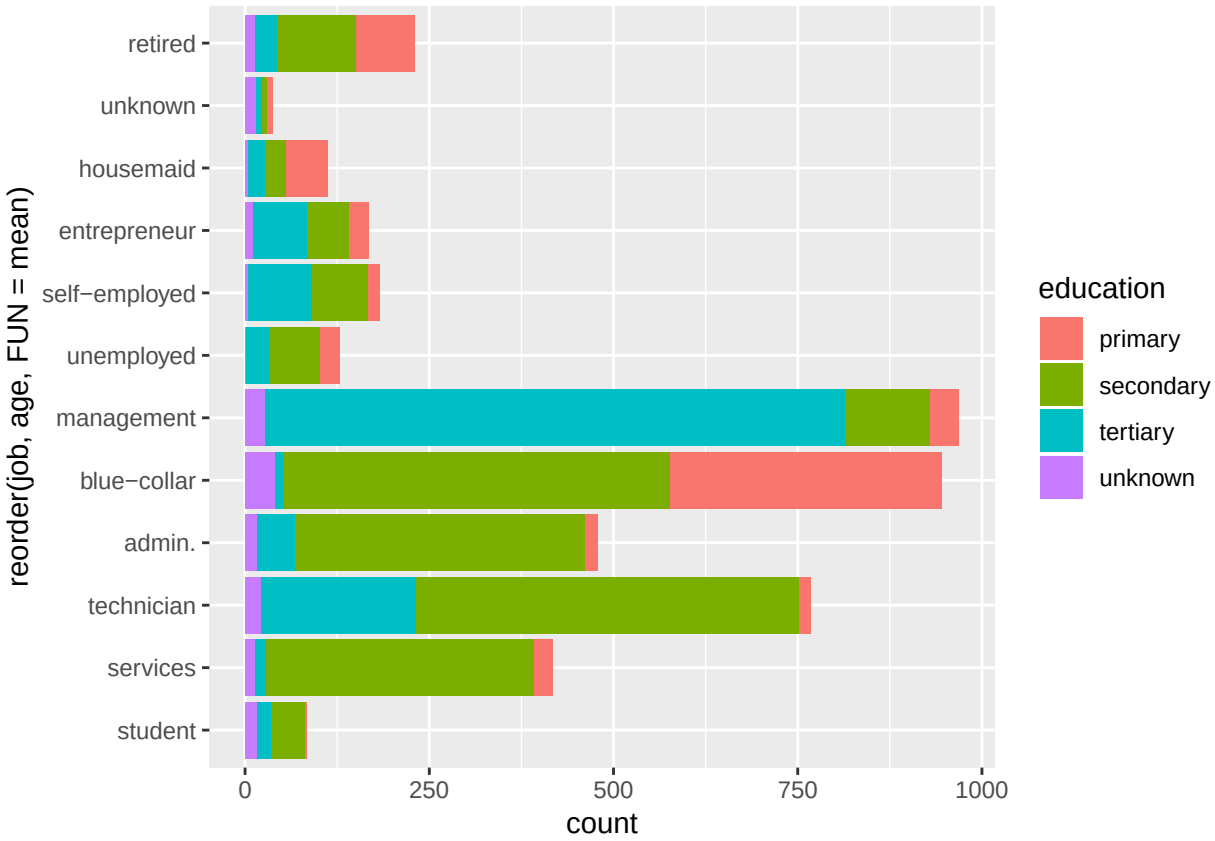
December 3, 2018

Data Analysis- Bank DataFrame. Following are the insights seen.

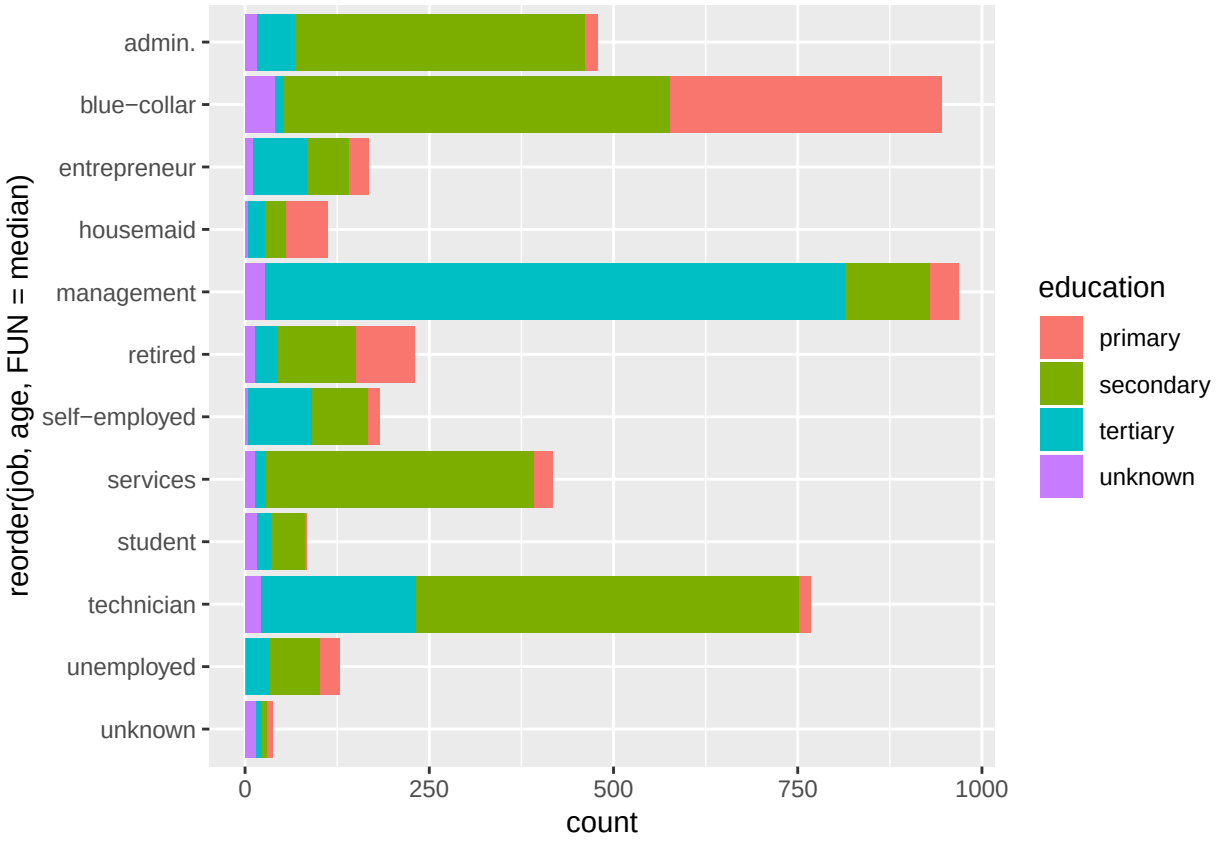
1. Sorting job by mean age and observing % of education background visually

Observation: I have sorted job based on mean age in first plot.Housemaid mean age next highest to retired.May be people end up in this profession if they couldn't fit in some productive career in their early years. Most of Management has higher education. Highest % of primary is in blue collar jobs. For admin,services,technician most of them have education higher than secondary.Some one aiming for this jobs should try to have min secodnary. Next plot sorts job by alphabetical order.This time I used median age for sorting

```
ggplot(data = bank, mapping = aes(x = reorder(job, age, FUN = mean),fill = education)) +
   geom_bar() +
   coord_flip()
```



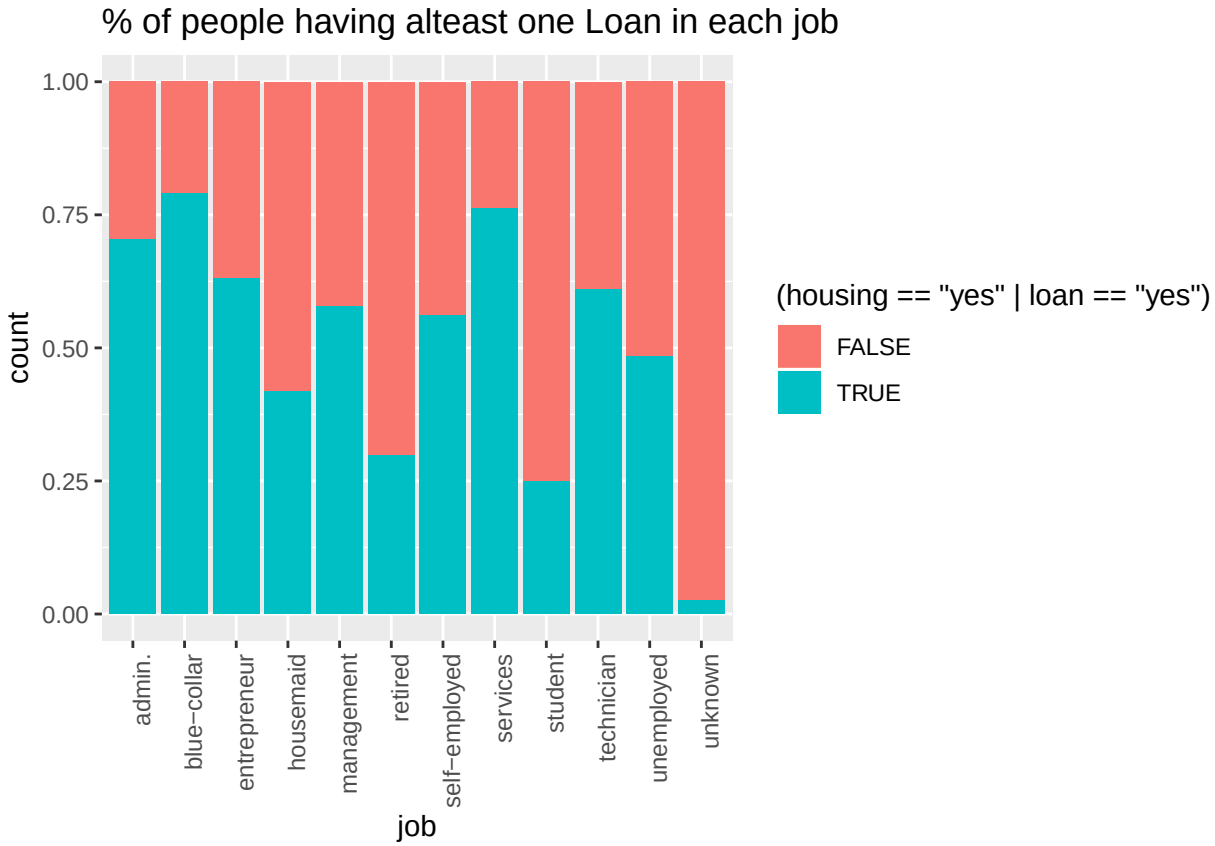

```
ggplot(data = bank, mapping = aes(x = reorder(job, age, FUN = median),fill = education)) +
  geom_bar() +
  scale_x_discrete(limit = rev(levels(bank$job))) +
  coord_flip()
```



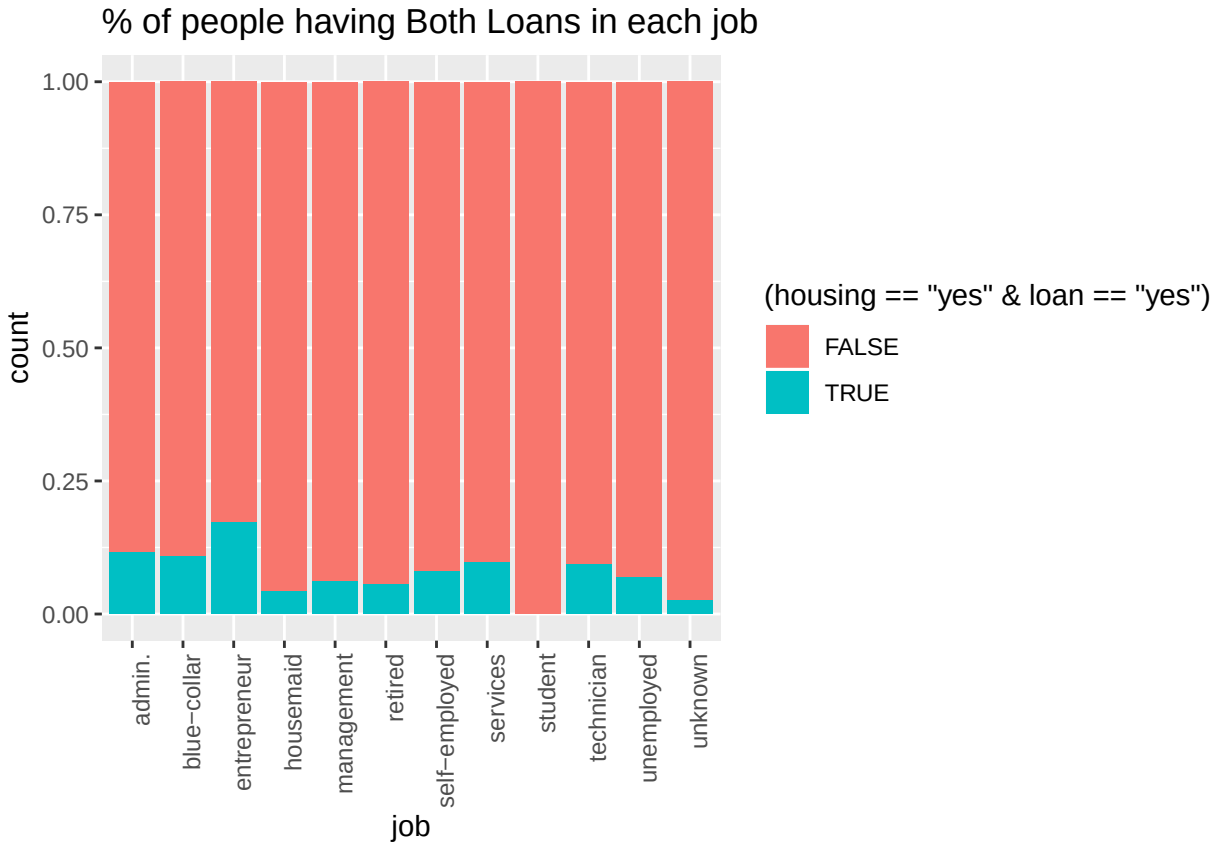
2. Analsying people with loans(both housing,personal or any one) based on job.

Observation: Drastic difference between both scenarios.very few people take both loans. First Scenario:Atleast one Loan Student,retired has least % of loans relatiev to other profession which is intuitive. Admin,Bluecolar,services,entreprenuer hass highest,close to 75% . Bank employees need to target people in this profession .Higher chances to open loan accounts. Second Scenario:Having both loans No student has both loans. There is highest % in Entreprenurs (having both loans) which is understandable given the risks they take.

```
ggplot(bank) +
  geom_bar(aes(x=job, fill=(housing=='yes' | loan=='yes')), position="fill") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_x_discrete(limit = levels(bank$job))+
  labs(title="% of people having alteast one Loan in each job")
```




```
ggplot(bank) +
  geom_bar(aes(x=job, fill=(housing=='yes' & loan=='yes')), position="fill") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_x_discrete(limit = levels(bank$job))+
  labs(title="% of people having Both Loans in each job ")
```

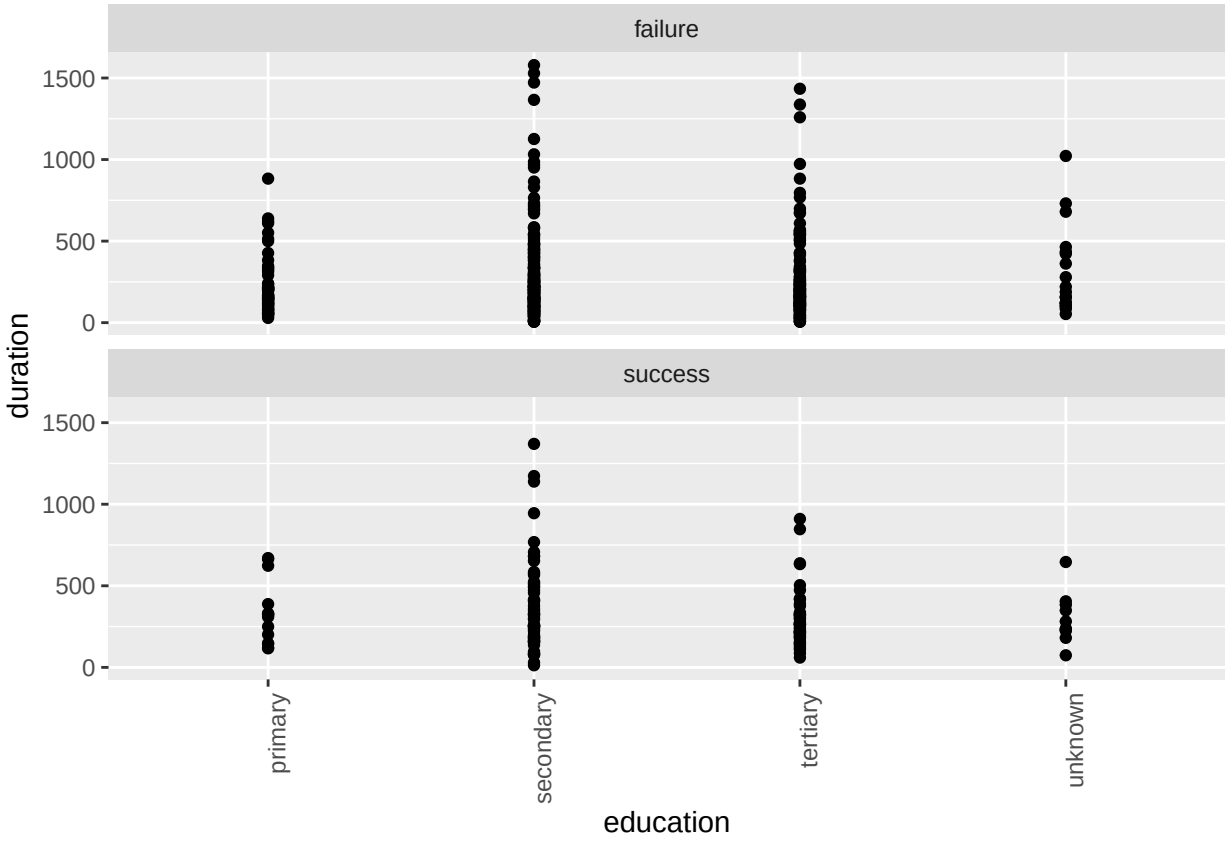


3. Trying to see any difference between duration of call for successful and failed outcome based on education.

Observation: Firstly people with higher education has more duration relative to primary.May be they have more doubts and answers to be received due to better knowledge they possess which is understandable. Duration call for successful outcome is less across education level. It's difficut to convince people and if duration is higher most likely it might end up being a faliure.

```
ggplot(data=bank%>% dplyr::filter(poutcome=="success"|poutcome=="failure"), aes(education, duration)) +
geom_smooth() +
facet_wrap(. ~ poutcome, nrow = 2) +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

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

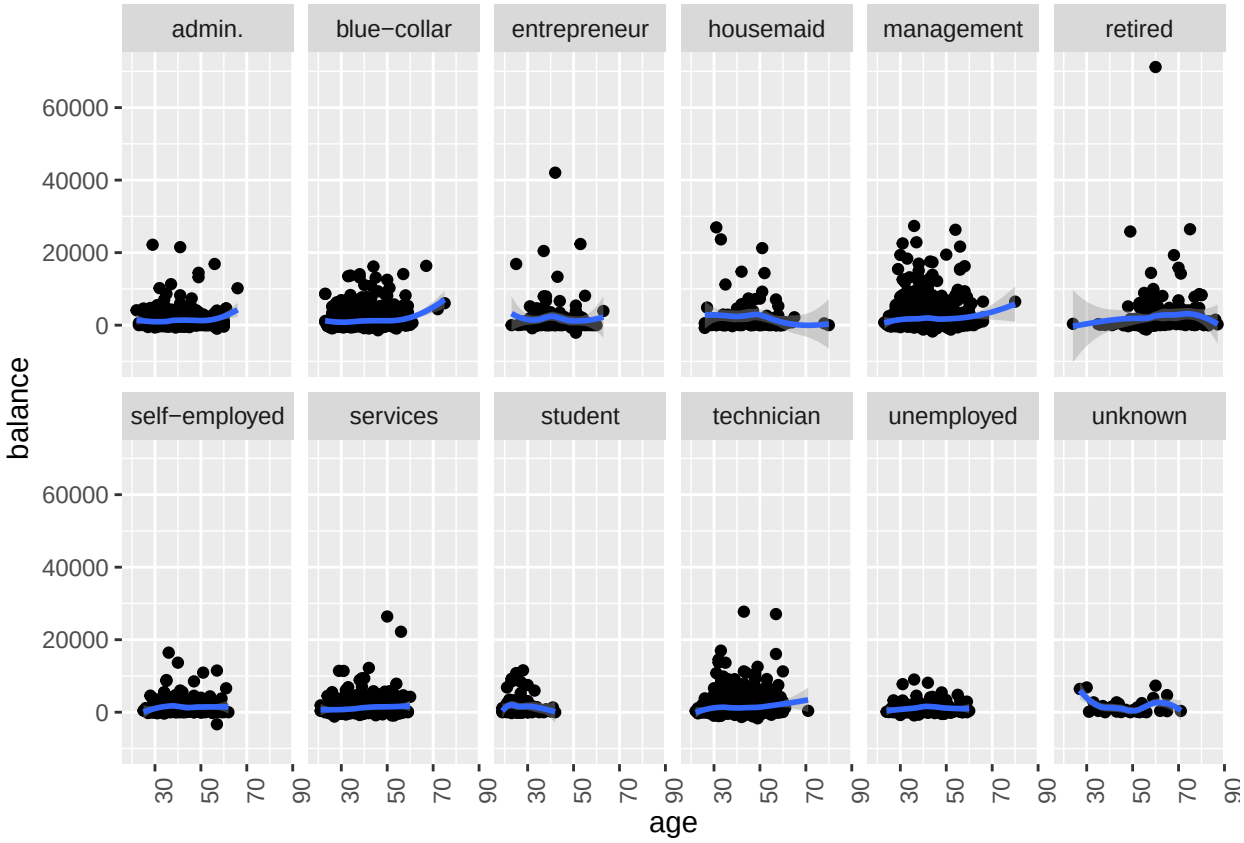


4. Balance wrt to age in each job.

Observation: Tried to use smooth line to see any trend in balance wrt to age Though there are few observations I don't find them worth to say it concretely. Observing blue line would give bit better change in balancee wrt to age.And also outliers are shown clearly in plot.

```
ggplot(bank, aes(age, balance)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(. ~ job, nrow = 2) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

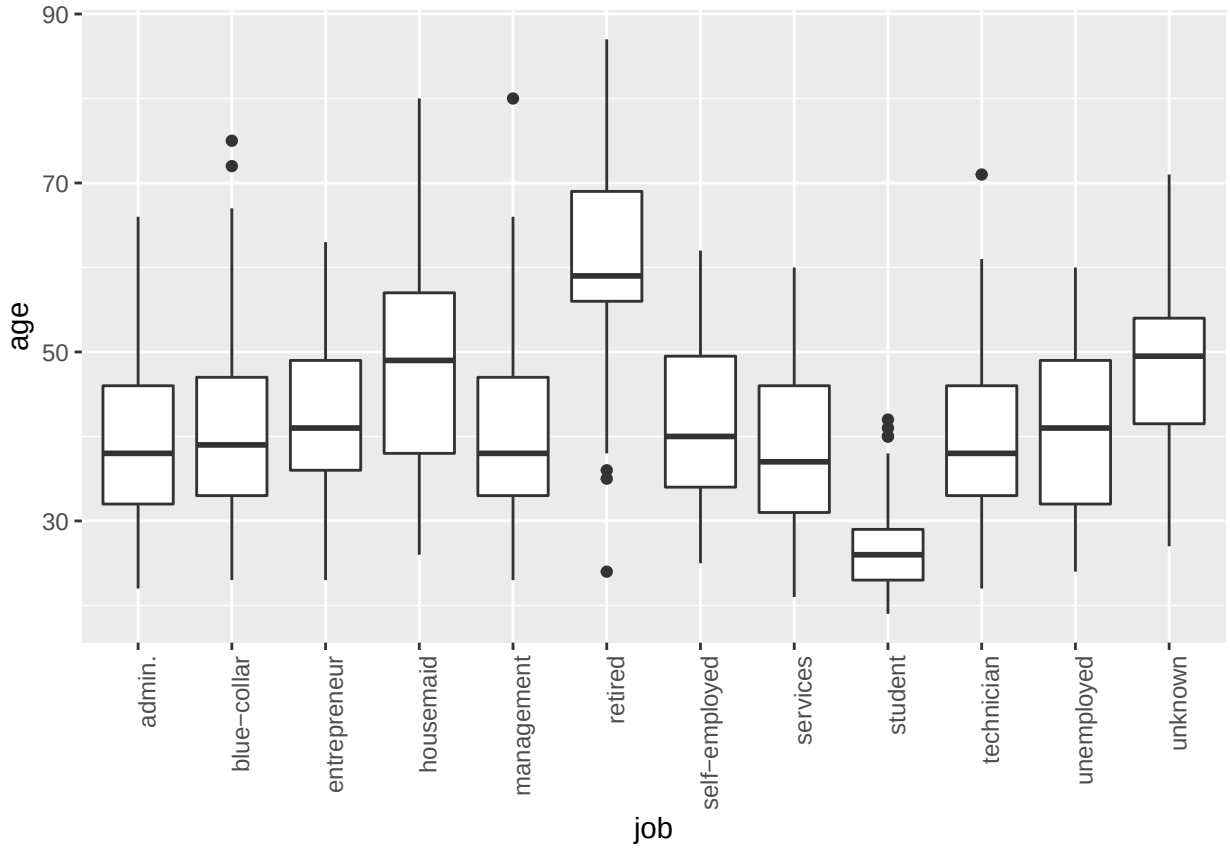
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



5. Box plot for age in each job

Observation: FEW OUTLIERS IN retired and in bluecollar,technician sounds interesting to me.People have retired at such a low age .People work as blue collar even in 75-80 age.

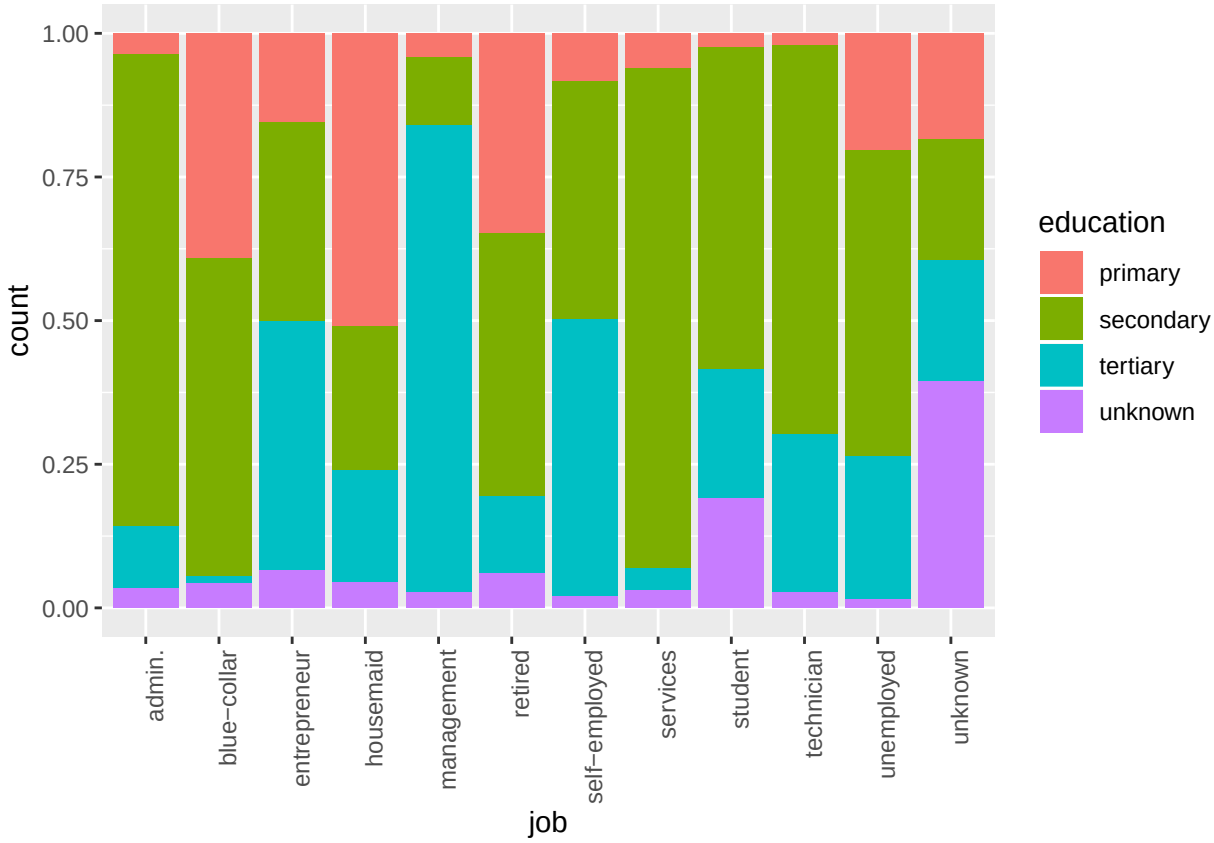
```
ggplot(bank, aes(job, age)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  geom_boxplot()
```



6. Education Background Stats based on job.

Observation: Highest % of higher eductaion in Management. Highest % of primary eductaion in housemaid. We ha dcome across with similar observations visually in finding 1.

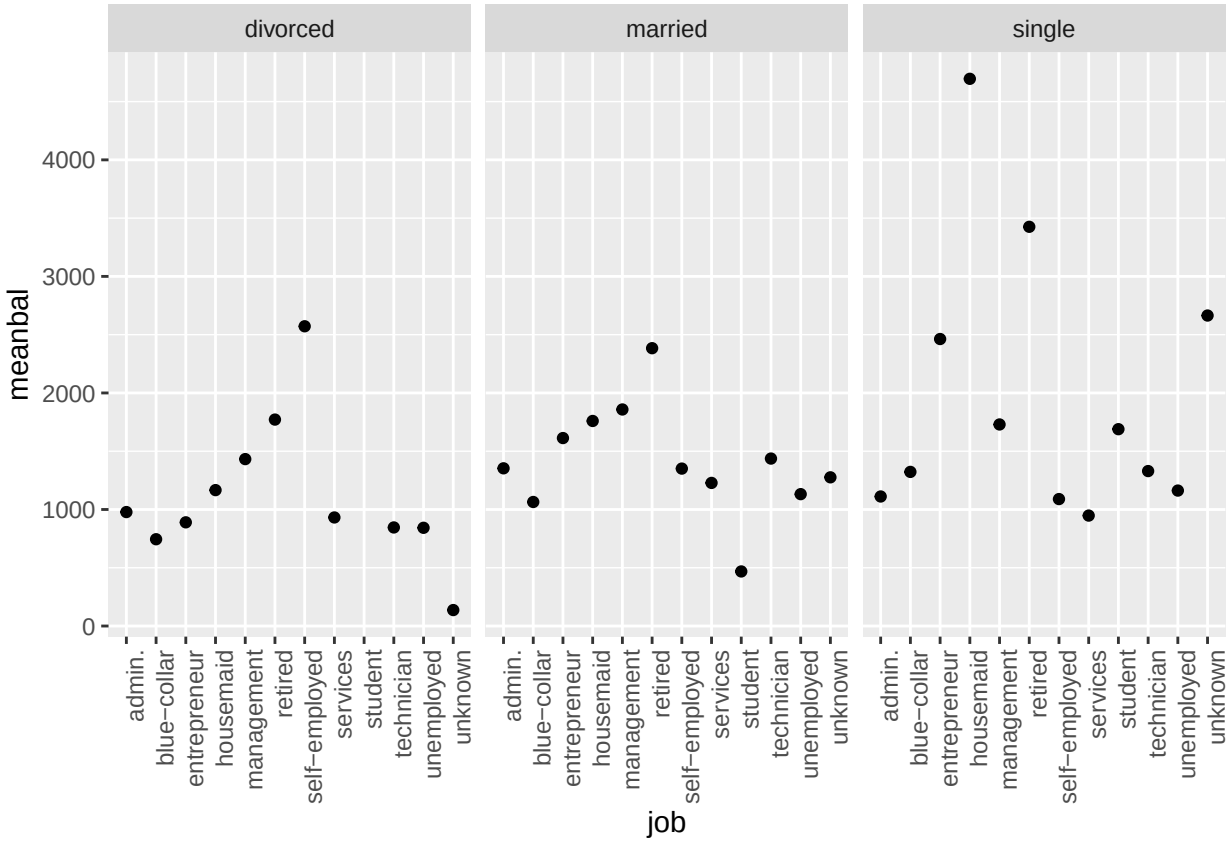
```
ggplot(bank) +geom_bar(mapping = aes(x = job, fill = education),position = "fill") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



7. Mean Balance wrt to marital status in each job

Observation: Huge differnce for House maid mena balance single Vs divorced Broadly, divorce people across all jobs has lower mean balance relatively and single across all jobs highest

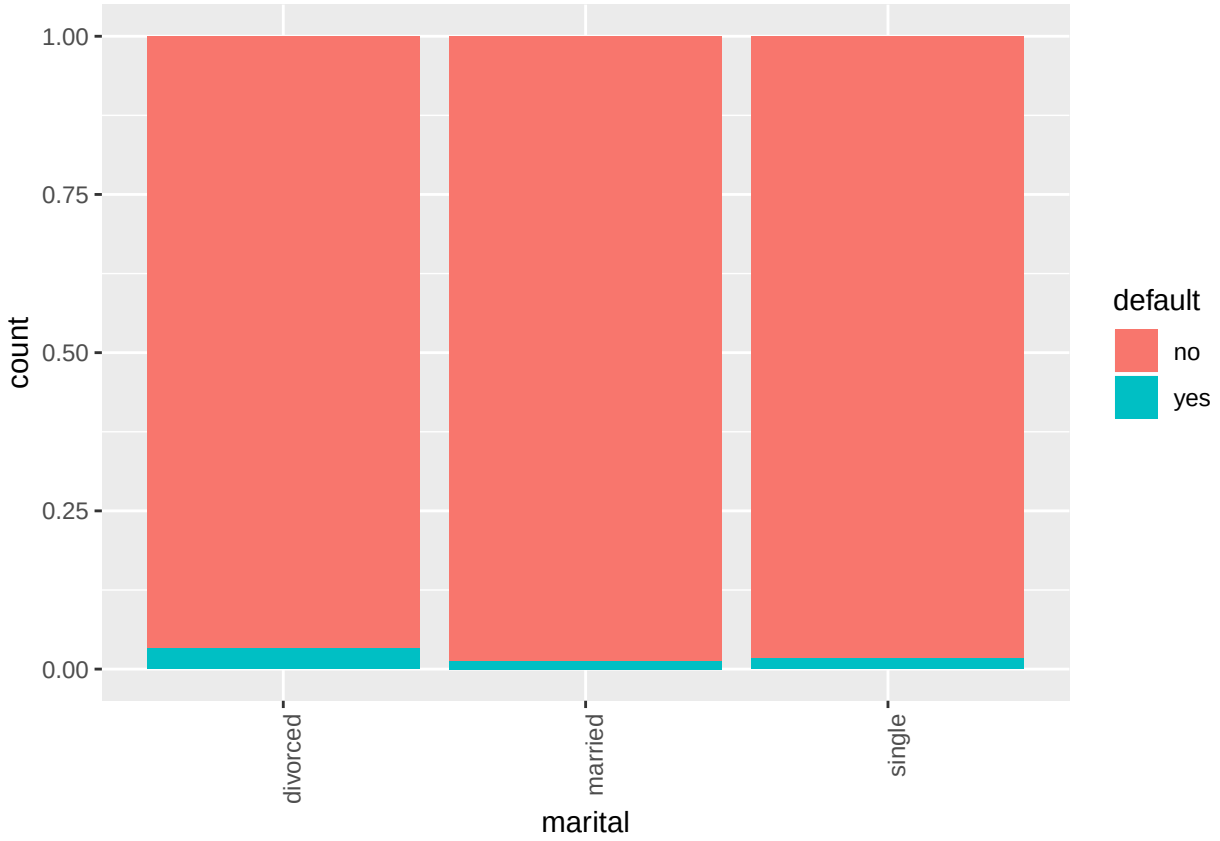
```
ggplot(bank) +geom_point(data=bank%>%group_by(job,marital)%>%summarise(meanbal=mean(balance)),mapping =
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



8. Default rate based on marital status

Observation: Insignificant factor. Defalut rates doesn't depend on marriage. For divorced people,relatively higher.

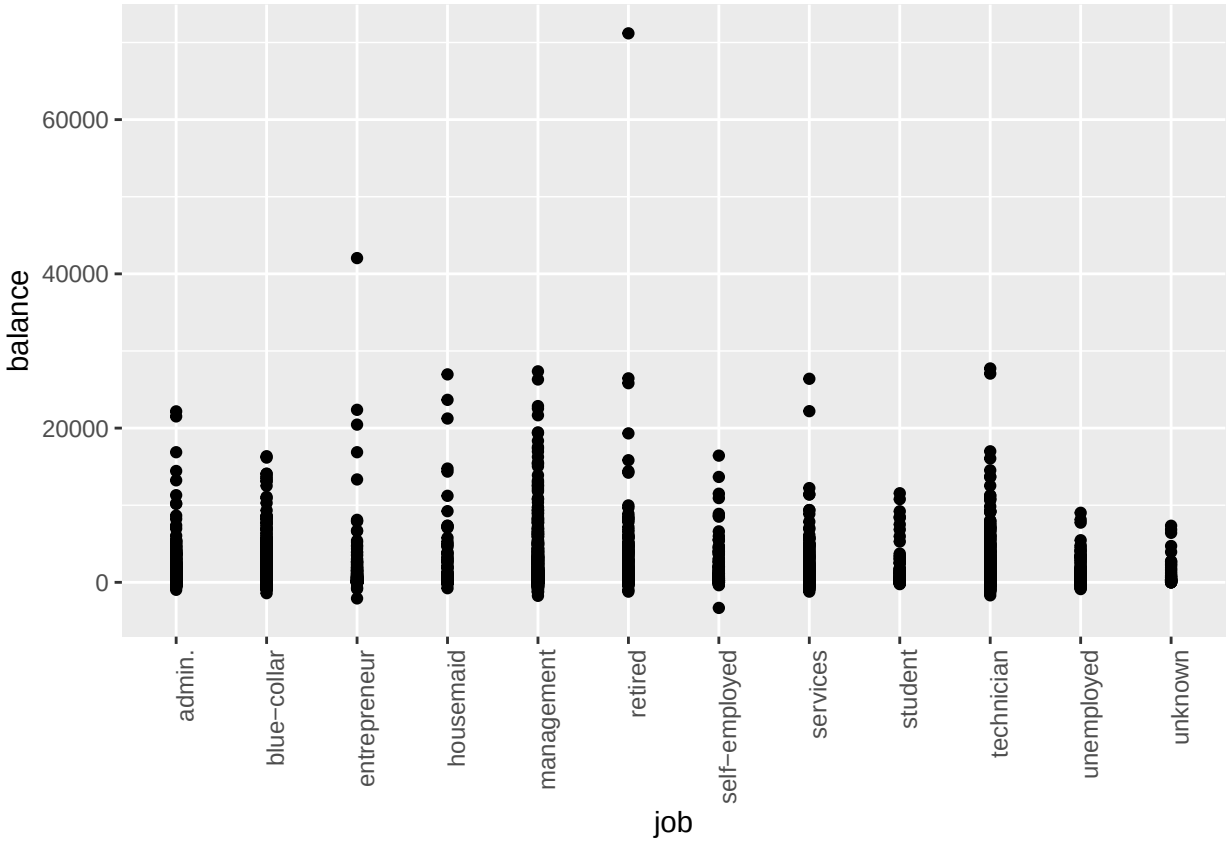
```
ggplot(bank) +
  geom_bar(mapping = aes(x = marital, fill = default),position = "fill") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



9. Balances of people across jobs

Observation: Overall I don't find anything significant difference.One particular profession earning so higher isn't there.Almost all the professions have negative balace. Some minute takeaways- relatively managament earns bit higher and surprisingly most of entreprenuers has less balance.Unemployed understandably has lowest.

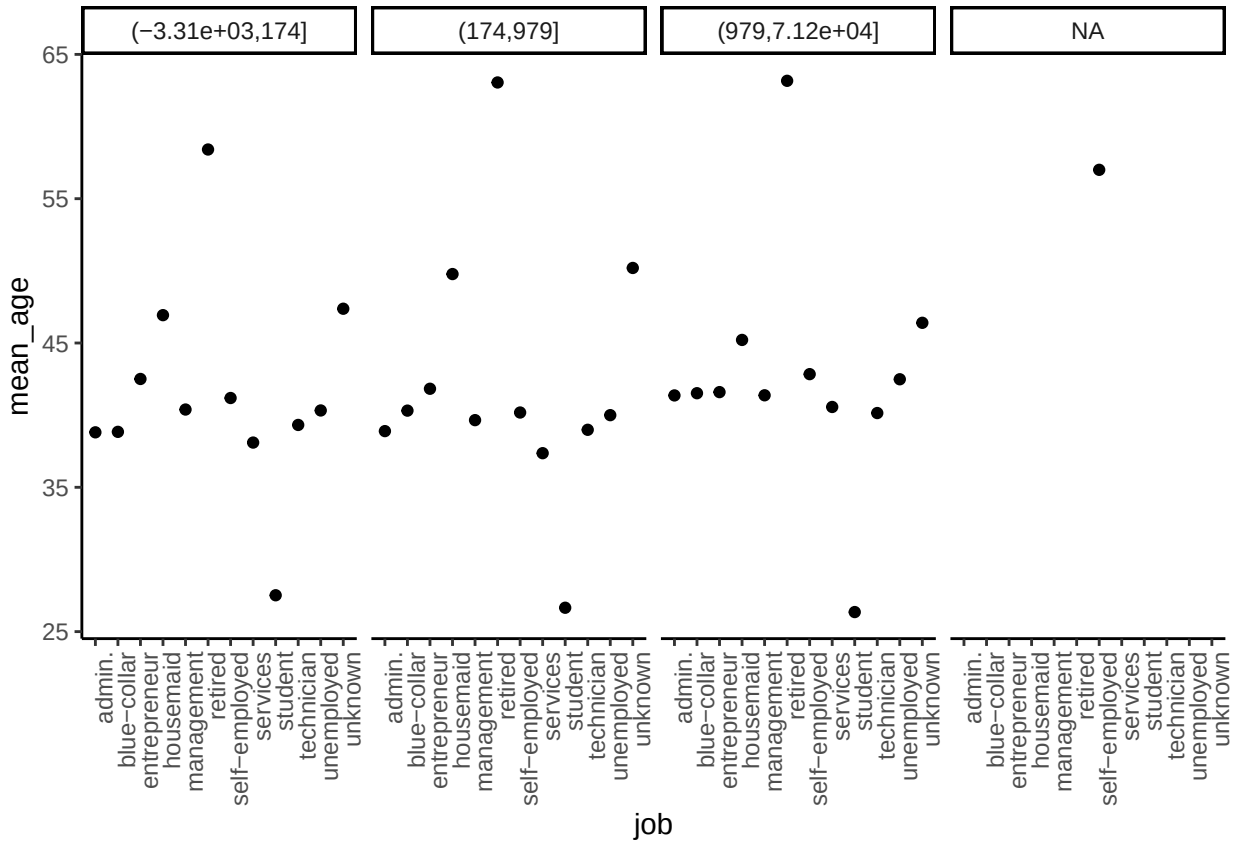
```
ggplot(bank) + geom_point(aes(job, balance)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



10. Divided data into 3 buckets based on balance.Plot mean age in each profession resp for each group

Observation: Plot shows well distributed i.e all this groups have similar mean age charactristics indicating all type of incomes are possible in any age.leaving few outleirs in each bracket,we can conclude that there isn't any siginificant relation in people balance wrt to age in all the professions.

```
cutpoints <- quantile(bank$balance, seq(0, 1, length = 4), na.rm = TRUE)
bank$balance_newgroup <- cut(bank$balance, cutpoints)
x<-bank%>%group_by(balance_newgroup,job)%>%summarise(mean_age=mean(age))
ggplot(x,aes(x=job,y=mean_age))+facet_wrap(~balance_newgroup,nrow=1)+geom_point()+theme_classic()+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```


(−3.31e+03,174]
(174,979]
(979,7.12e+04]
NA
mean_age
65
55
45
35
25
admin.
blue−collar
entrepreneur
housemaid
management
retired
self−employed
services
student
technician
unemployed
unknown
job