

# LivDivDataPlotsRD

Siddarth Ravikanti

6/8/2022

#Initial look at Livdiv data set and creating basic vizualizations #Import libraries needed to create plots

```
inc <-baseline %>% select(contains("inc"))
```

**Number of Children per household**

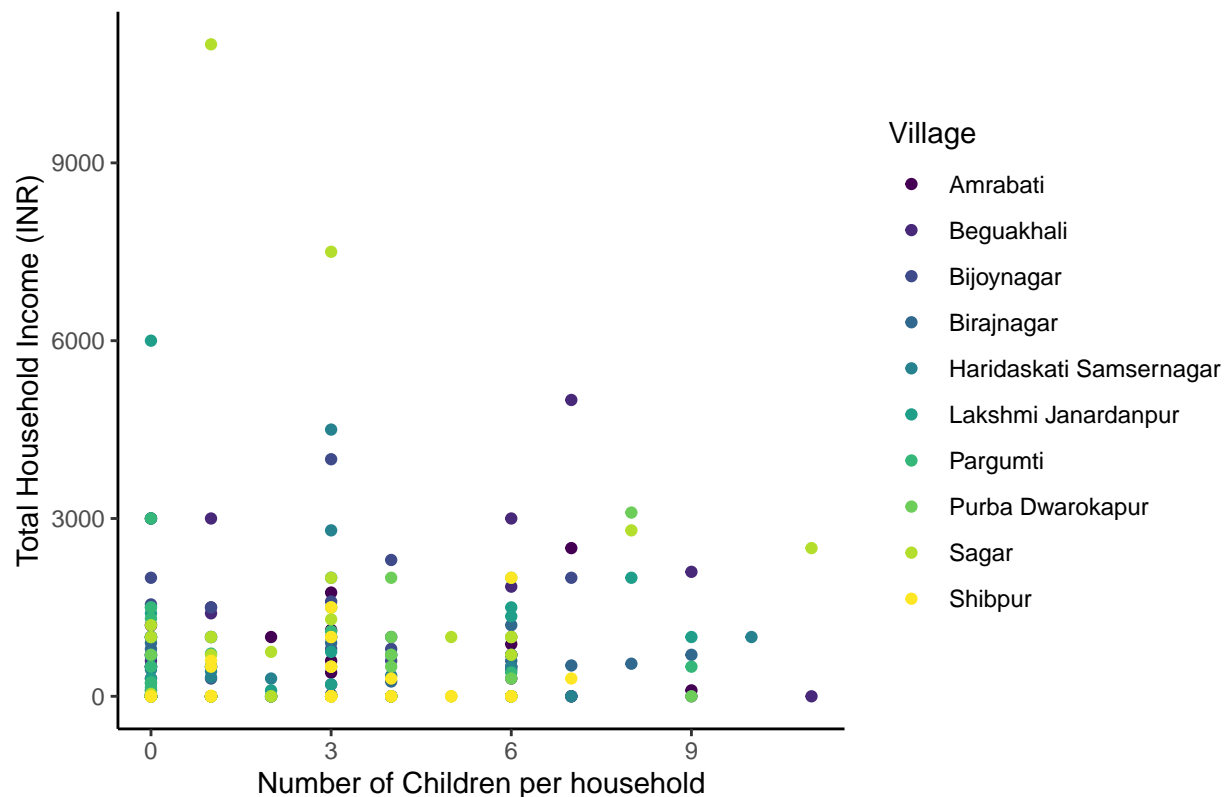
**scatter plot illustrating the number of children vs total household income**

**Analyzing to see if there is any correlation between both data points**

```
ggplot(baseline, aes(nb_children, inc_total, color = village)) + geom_point() + theme_classic() + labs()
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

Children per Household vs Household Income



```
summary(baseline$nb_children)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   0.00    2.00    2.52   4.00   11.00
```

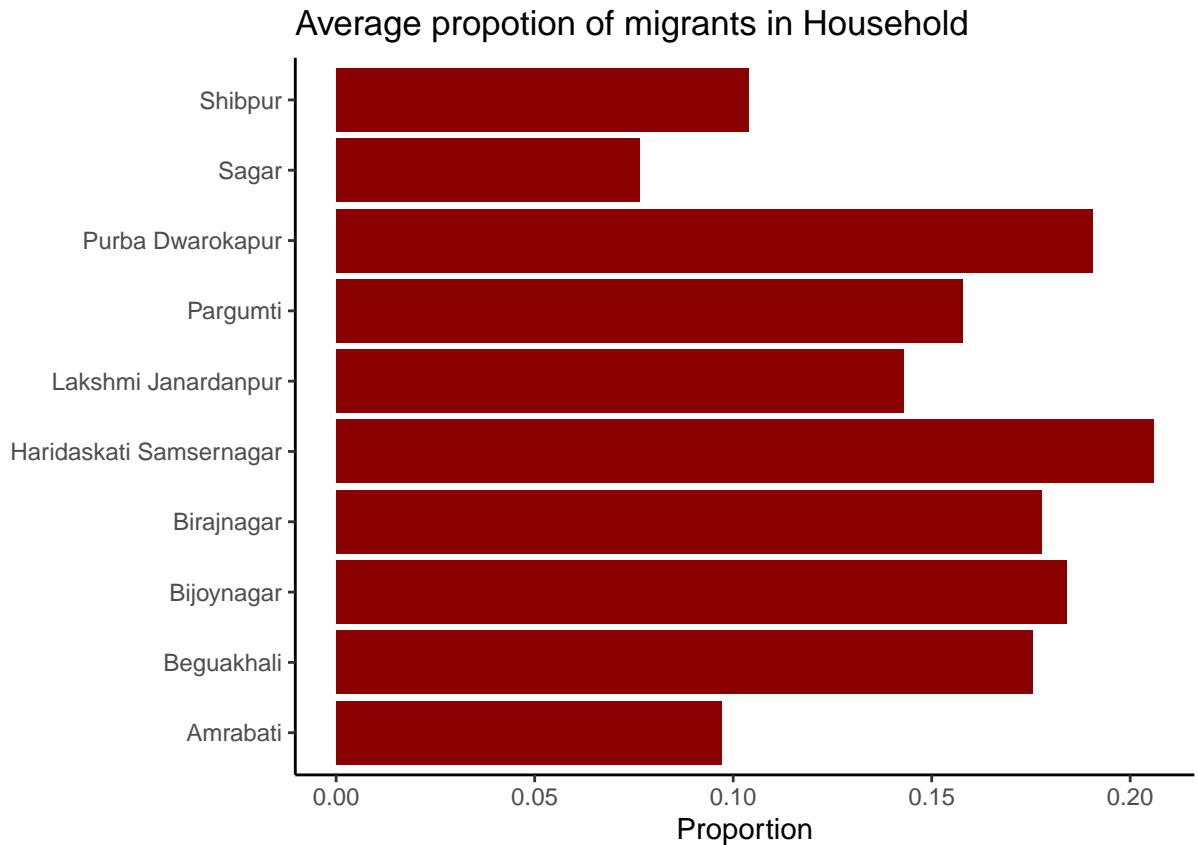
number children by household per village total household income per village

#What does the data potentially show? *The scatter plot indicates that the data set is somewhat bi modal with peaks at 0 children and 6 children. The median number of children is 2 and the mean is 2.52. The max number of children is 11. This can indicate that households are having less children due to the cost of raising a child This can also indicate that some households still choose to have a lot of children due to social norms in rural Indian communities* \*Poorer women tend to have more children, not only because wealth is connected with access to education and health care, but also because children can contribute labor and earnings to a household with limited means.

\*Source - <https://www.pewresearch.org/religion/2021/09/21/causes-of-religious-change/>

##“Proportion of HH with migrant”

```
migrant_prop <- baseline %>% group_by(village) %>% summarize(migrant_proportion = mean(hh_migrant_prop))
ggplot(migrant_prop, aes(village, migrant_proportion)) + geom_col(fill = "dark red") + theme_classic()
```



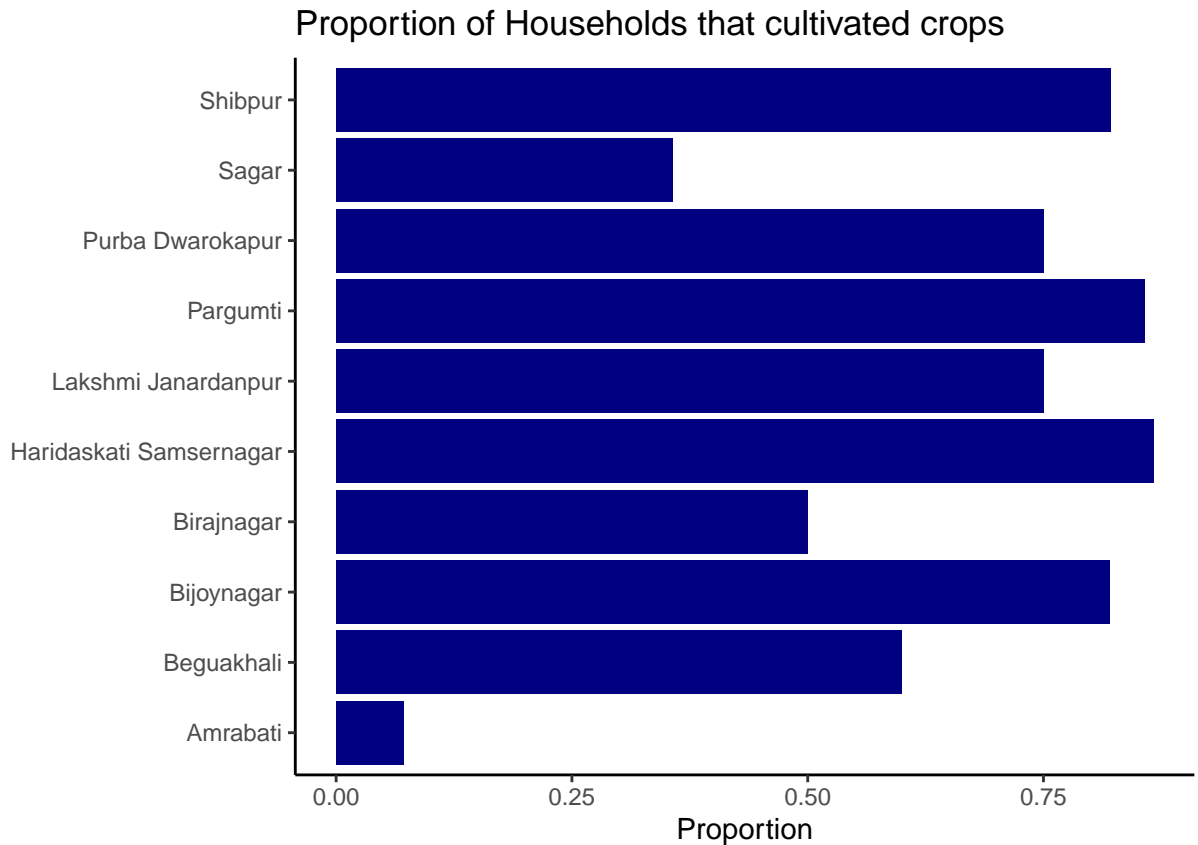
```
summary(migrant_prop$migrant_proportion)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.07653 0.11372 0.16660 0.15118 0.18242 0.20587
```

What does the data potentially show? \* Majority of the HH don't have any migrants in the household \* The village of Sagar has the lowest avg migrant population per household \* This can indicate that the households in the village receive the lowest amount of remittance \* This can also indicate that these households rely on the income sources around them \* On average, around 15% of the household can be classified as migrant \* Summary data \* Median proportion of household migrants is .1667 \* Mean proportion of household migrants is .1541

```
## "Proportion of HH involved in agriculture"
```

```
grouped <- baseline %>% group_by(village) %>% summarize(prop_farm = sum(farm_yn)/n())
ggplot(grouped, aes(village, prop_farm)) + geom_col(fill = "navy blue") + labs(x = "", y = "Proportion",
```



```
summary(grouped$prop_farm)
```

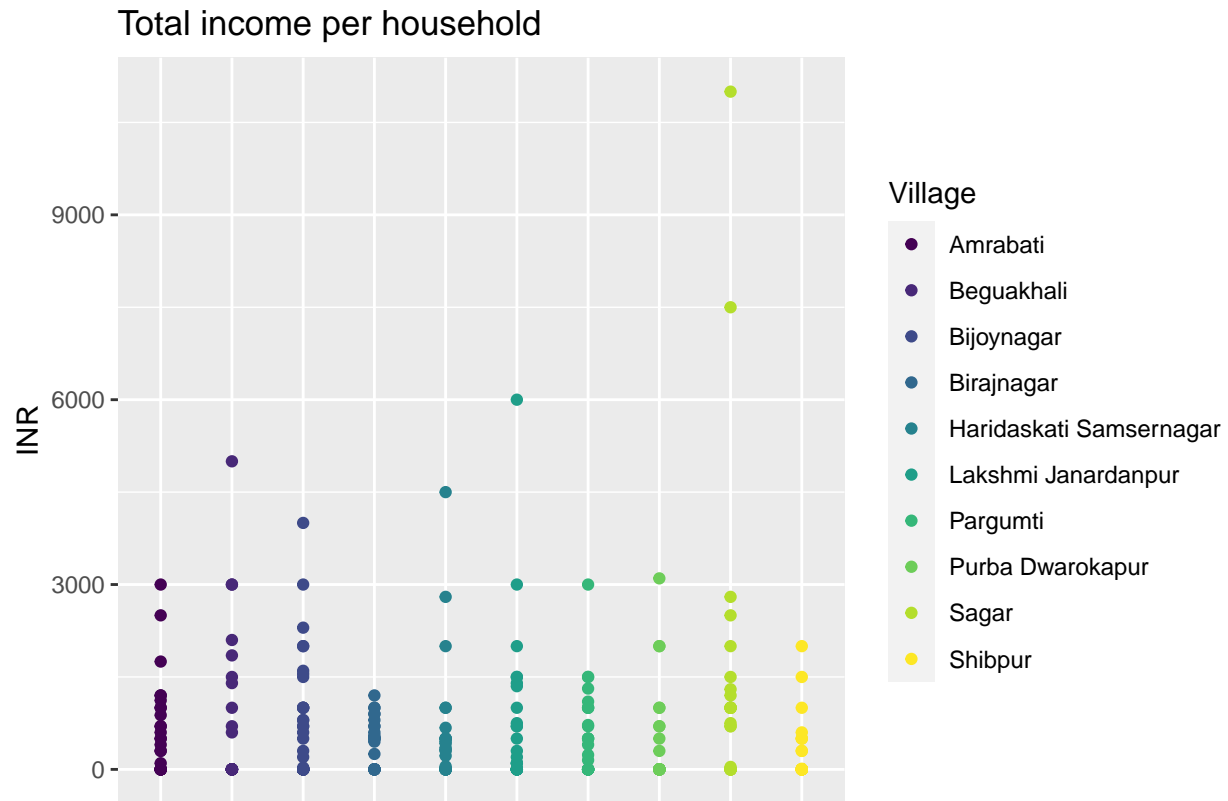
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.07143 0.52500 0.75000 0.63938 0.82107 0.86667
```

#What does the data potentially show? *Amrabati by far has the lowest proportion of households that participate in agriculture.* So does this mean that this village has a lower or higher avg household income than the other villages or is the opposite. \* What other sources income do they have? \* On average, across all the villages, 63.9% of the households participate in farming with the median being 75%. \* This is important because almost 3/4 of the households are dependent on farming as a way of life. \* Important followup questions can be if education level and working in the farm

## Total Household Income

```
ggplot(baseline, aes(inc_total, village, color = village)) + geom_point() + labs(x = "INR", y = "", tit
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
summary(baseline$inc_total)
```

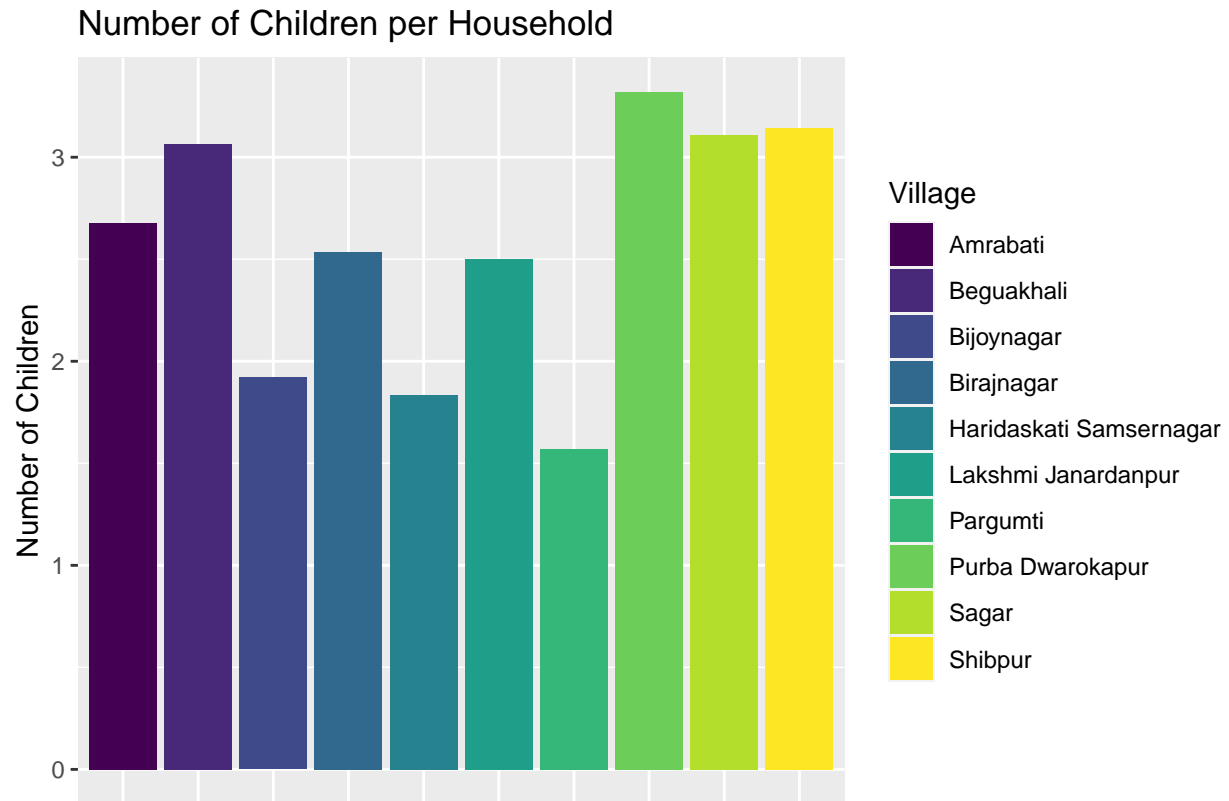
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.      NA's
##      0.0    0.0    10.0   606.4   880.0 11000.0         1
```

##Summary of the statistics #From the baseline statistics, the mean total income for all the villages is 606.4 INR #The max total income of a household is 11,000 INR coming from a household in Sagar. # 1 NAs were reported and removed from histogram

#Number of children per household

```
avg_children <- baseline %>% group_by(village) %>% summarize(avg_children = sum(nb_children)/n())
```

```
ggplot(avg_children, aes(village, avg_children, fill = village)) + geom_col() + labs(x = "", y = "Number of children per household")
```



#Agricultural landholding

```
land <- baseline %>% select(village, no_farm_reason) %>% na.omit(no_farm_reason)
```

```
land$no_farm_reason <- as.numeric(as.factor(land$no_farm_reason))
```

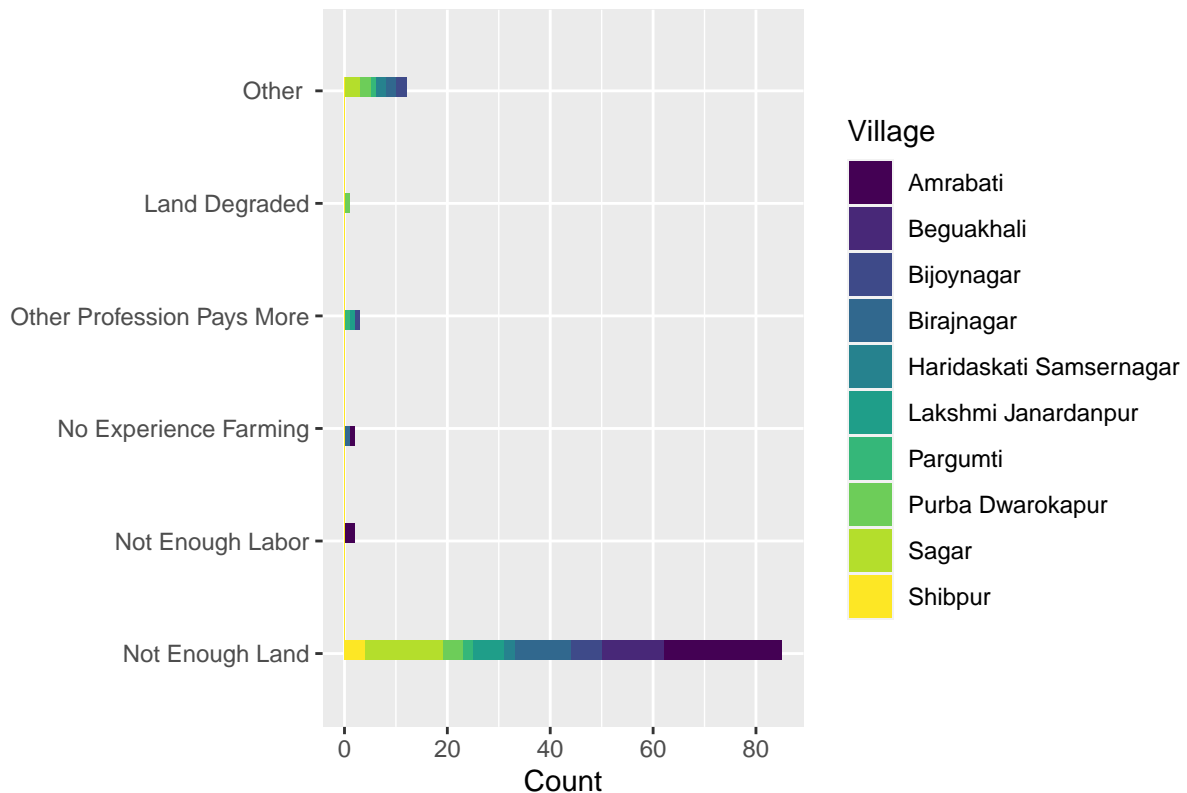
```
ggplot(land, aes(no_farm_reason, fill = village)) + geom_histogram() + labs(x = "", y = "Count", title = "Number of Children per Household")
```

```
## Warning: Continuous limits supplied to discrete scale.
```

```
## Did you mean 'limits = factor(...)' or 'scale_*_continuous()'?
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

## Reasons Household Doesn't Farm



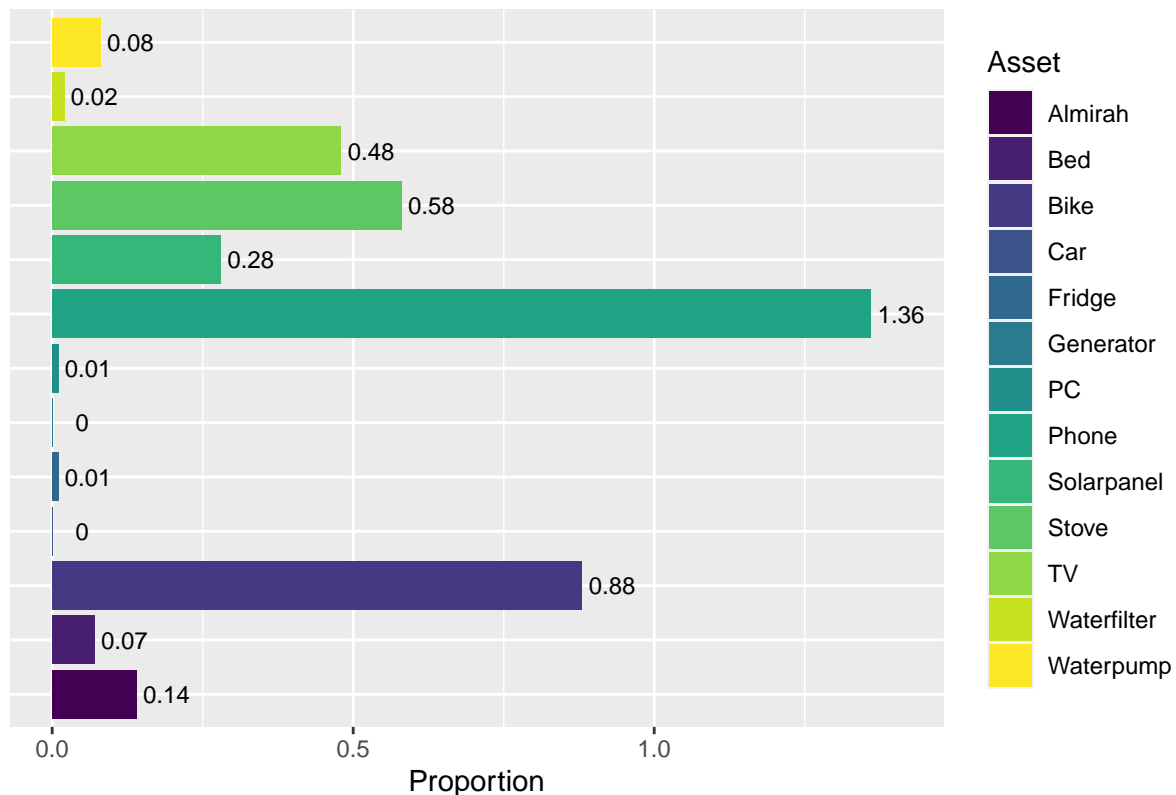
#What does the data potentially show? This data shows that the biggest reason why a household will not partake in agriculture is because there is not enough land. Land degradation is the lowest reason why the households are not taking part in agriculture. Amrabati specifically has the most reasons why the households there don't take part in agriculture, meaning that this issue is more prevalent in that village. Not enough land can mean that either they don't have access to plots of land to grow crops on or the land itself is not arable.

#Household Assets (by type)

```
assets <- baseline %>% select(contains("asset")) %>% select(contains("num")) %>% summarize(Stove = sum(
assets_long <- gather(assets, property, measurement, Stove:Waterfilter)
assets_long["measurement"] <- round(assets_long$measurement, digits = 2)

ggplot(assets_long, aes(property, measurement, fill = property)) + geom_col() + labs(x = "", y = "Propor
```

## Proportion of Households Owning Assets



#What does the data potentially show? Mobile phones are the most common asset owned by these households. The lowest owned asset are cars as well as solar panels and generators. Expensive assets like cars, generators and etc. are rarely seen among the households. \*Almirah's are a type of wardrobe/storage unit

Important crops

```
baseline$imp_crop1 <- as.numeric(as.factor(baseline$imp_crop1))

prop_crops1 <- baseline %>% select(village, imp_crop1) %>% filter(imp_crop1 == 1) %>%
  summarize(crop1 = sum(imp_crop1)/count(baseline))
prop_crops2 <- baseline %>% select(village, imp_crop1) %>% filter(imp_crop1 == 2) %>%
  summarize(crop2 = (sum(imp_crop1)/2)/count(baseline))
prop_crops3 <- baseline %>% select(village, imp_crop1) %>% filter(imp_crop1 == 3) %>%
  summarize(crop2 = (sum(imp_crop1)/3)/count(baseline))
prop_crops4 <- baseline %>% select(village, imp_crop1) %>% filter(imp_crop1 == 4) %>%
  summarize(crop2 = (sum(imp_crop1)/4)/count(baseline))
prop_crops5 <- baseline %>% select(village, imp_crop1) %>% filter(imp_crop1 == 5) %>%
  summarize(crop2 = (sum(imp_crop1)/5)/count(baseline))
prop_crops6 <- baseline %>% select(village, imp_crop1) %>% filter(imp_crop1 == 6) %>%
  summarize(crop2 = (sum(imp_crop1)/6)/count(baseline))

important_crop <- data.frame(c(prop_crops1,prop_crops2,prop_crops3,prop_crops4,prop_crops5,prop_crops6))

important_crop_long <- gather(important_crop, crop, measurement, n:n.5)

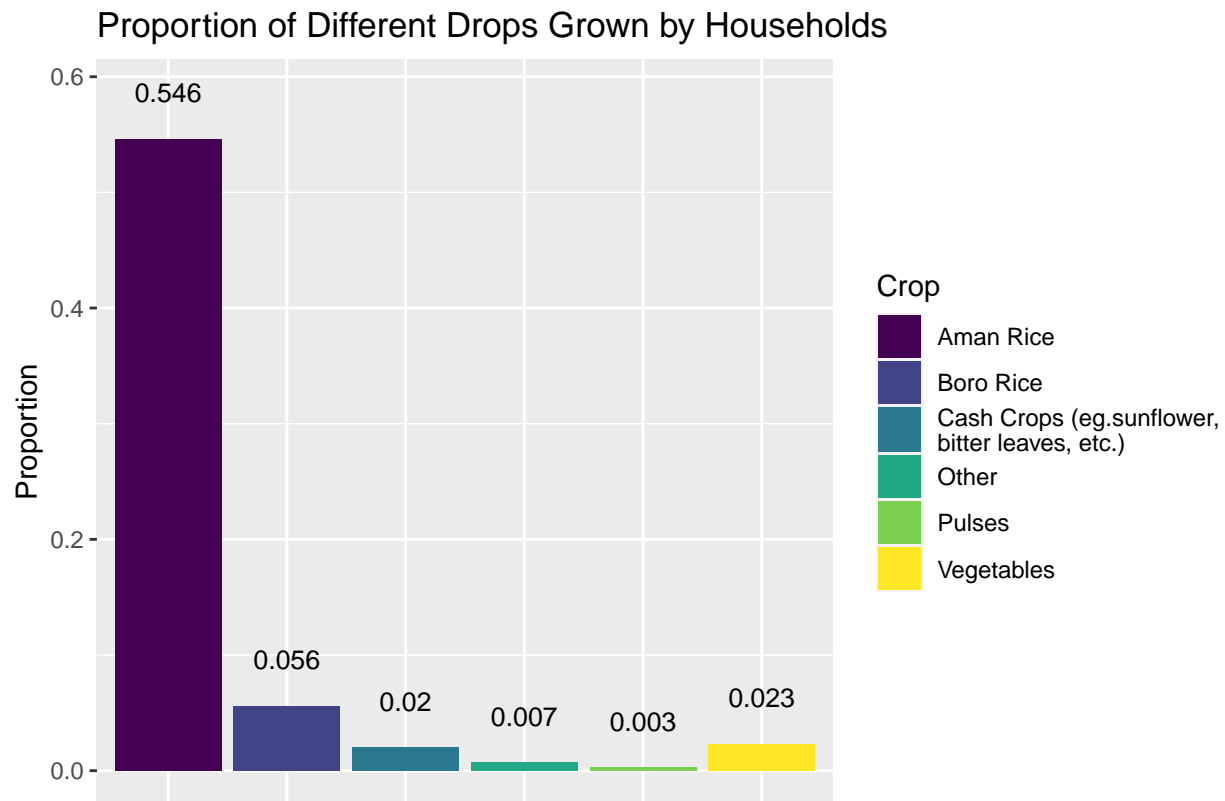
important_crop_long$crop <- c("Boro Rice", "Aman Rice", "Vegetables", "Cash Crops (eg.sunflower,
```



```
bitter leaves, etc.)", "Pulses", "Other")

important_crop_long["measurement"] <- round(important_crop_long$measurement, digits = 3)

ggplot(important_crop_long, aes(crop, measurement, fill = crop)) + geom_col() + labs(x = "", y = "Proportion")
```

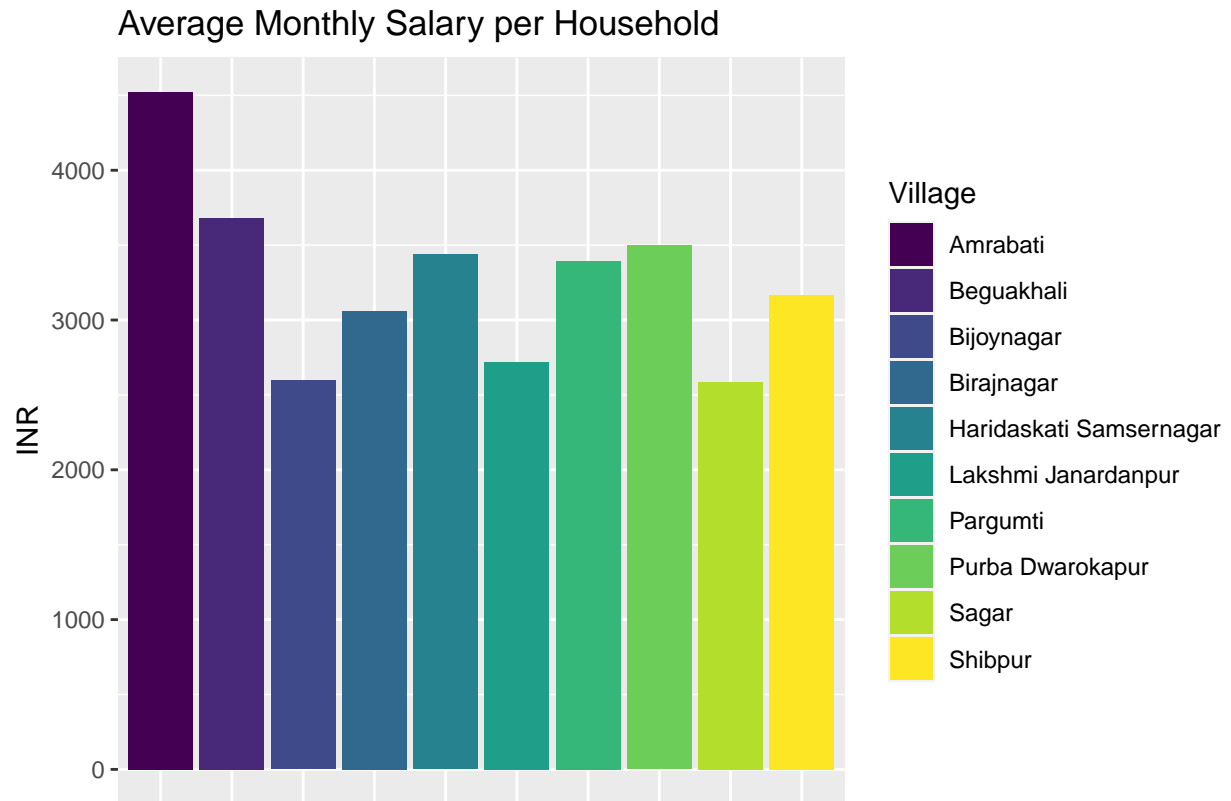


#What does the data potentially show? Aman rice is the most grown crop by the households. Pulses and other crops are the lowest grown. Monthly Salary

```
m_salary <- baseline %>% group_by(village) %>% select(job1_salary1) %>% summarize(avg_salary = sum(job1_salary1))
```

## Adding missing grouping variables: 'village'

```
ggplot(m_salary, aes(village, avg_salary, fill = village)) + geom_col() + labs(x = "", y = "INR", title = "Monthly Salary by Village")
```



#What does the data potentially show? The data shows that the monthly salaries for the households that receive then range anywhere between 2500 INR to 4600 INR per month. On average, Amrabati receives the highest income and Sagar has the lowest.