# Project Proposal

*Shyam BV*

*October 15, 2016*

```
# load data

loandata <- read.csv ("./data/LoanStats_2016Q1.csv",header = TRUE,skip = 1, sep = ",",stringsAsFactors
```

**Research question**

**You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.**

Below are some research questions from Sample data

1. What is the mean funded loan amount different purpose and comparision?
2. Which state has highest loan request?
3. What is the range or interval of Lending loan for one quarter?
4. What will be the revenue for next month or quarter?
5. How many loans requests will Lending loan get for next month?

**Cases**

**What are the cases, and how many are there?**

For 2016 Q1, there are around 133K observations. We might need some sample data from it.

**Data collection**

**Describe the method of data collection.**

Data is collected from Lending loan website (https://www.lendingclub.com/info/download-data.action).

**Type of study**

**What type of study is this (observational/experiment)?**

This is an observational study.

**Data Source**

**If you collected the data, state self-collected. If not, provide a citation/link.**

Link: https://www.lendingclub.com/info/download-data.action

**Response**

**What is the response variable, and what type is it (numerical/categorical)?**

In all conditionals the response variable is provided loan amount to a person. It is an numerical data.

**Explanatory**

**What is the explanatory variable, and what type is it (numerical/categorical)?**

Explanatory variable are state, purpose, loan quantity.

**Relevant summary statistics**

**Provide relevant to your research question. For example, if you're comparing means across groups provide means, SDs, sample sizes of each group. This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.**

Sample size = 1000

```r
loandatamean <- group_by(loandata,purpose) %>% summarise(mean(funded_amnt))

loandatasd <- group_by(loandata,purpose) %>% summarise(sd(funded_amnt))
```