How do different farm characteristics impact the quantity of debt it has?

BY LINE O., JULIETTE P., PERRINE L. & NAZAR R.

Introduction

Objective of analysis: To understand the key factors influencing farm debts.

Focus Areas:

- Types of farms most affected by debt.
- Role of farm characteristics like size, production type, and location in financial burdens.
- Impact of farmer demographics, such as age and protected area status, on debt levels.

Significance:

- Rising farm debts in France: From 50,000 € in 1980 to 187,000 € in 2017.
- 22% of farmers lived below the poverty line in 2017.

(viande.info, n.d.)

Data

- The Réseau d'Information Comptable Agricole (RICA) is a survey realized by each member state of the European Union every year since 1968. It displays a variety of characteristics (quantitative and qualitative) to better assess the diversity of incomes in the sector.
- To make our calculations more manageable we decided to focus our work on certain rows and columns of the dataset:

ROWS

- Production type of the farm : choose in function of which has the largest amount of debt

COLUMNS

- Different farm characteristics which may impact the quantity of debt

Data

Selection of Farm Production Type (Technico-economic orientation in 17 positions: OTEXE) to focus on:

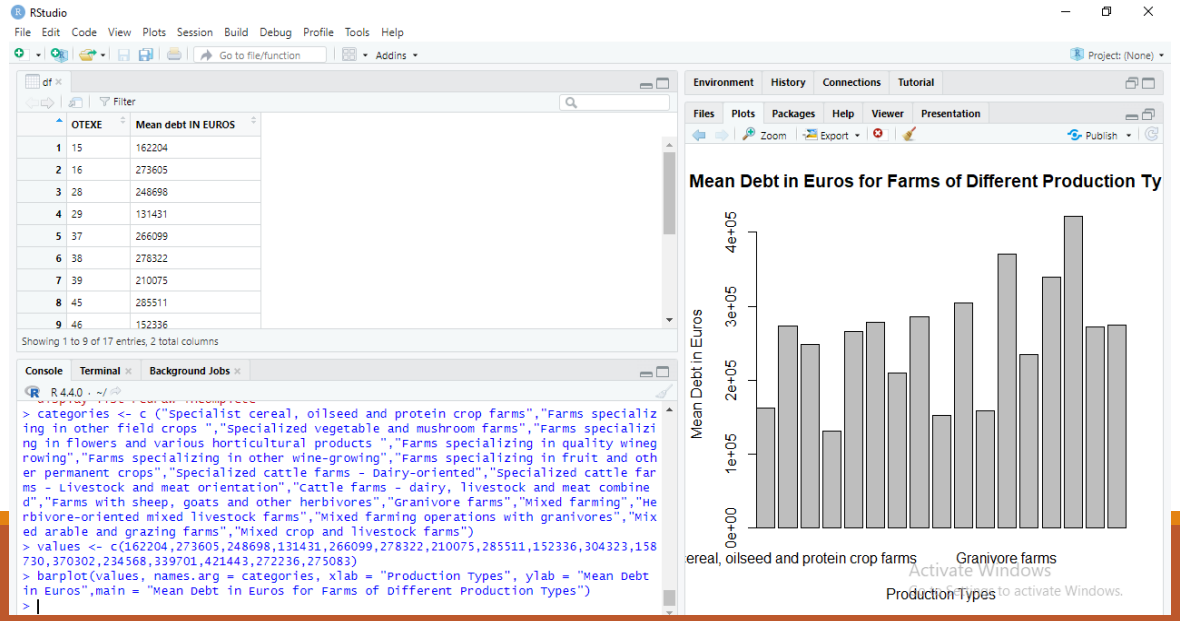
1) Exporting data

```
> fn="/Users/lenovo/Documents/TypevsDebt.csv"
> df = read.csv(fn, header = TRUE, sep = ",", dec = ".")
```

2) Filtering each type to find the mean of each 17 production type

```
> summary(df[df$OTEXE..Ptype. == "15",])
 OTEXE..Ptype.       Debt
 Min.   :15    Min.   :      0
 1st Qu.:15    1st Qu.:  42243
 Median :15    Median : 111764
 Mean   :15    Mean   : 162204
 3rd Qu.:15    3rd Qu.: 229248
 Max.   :15    Max.   :1171525
```

3) With all the calculated mean debt values constructing a bar chart:



Final descision:

Largest mean debt was found on the farm production type 74; “**Mixed farming operations with granivores**”.

Data

Selection of different factors to investigate :

Quantitative :

- total area of the farm (SUTOT)
- revenues of the farm(CHAFF)
- Debt quantity of the farm (TDTE3)

Qualitative:

- Age range of the farmer manager (TRA05)
- Sex of the farmers (SEXEP)
- Is the farm present on a Natura 2000 Zone (ZENVI)

Questions

1. Does the farm's size (in terms of surface area and revenue) impact the quantity of debt the farm has?
2. Does the sex of the farmer impact the quantity of debt the farmer has?
3. Does the age of the farmer impact the quantity of debt the farmer has?
4. Do farms present on natura 2000 zones have more debt?

1) Does the farm's size (in terms of surface area and revenue) impact the quantity of debt the "**Mixed farming operations with granivores**" farms have?

- Chosen test: Linear Regression Test
- Explanatory variables (x variables) = surface area (surface unit) and revenues (in euros)

SUTOT	1	Surface non nulle et inférieure à 5 ha
SUTOT	2	Surface égale ou supérieure à 5 ha et inférieure à 10 ha
SUTOT	3	Surface égale ou supérieure à 10 ha et inférieure à 15 ha
SUTOT	4	Surface égale ou supérieure à 15 ha et inférieure à 20 ha
SUTOT	5	Surface égale ou supérieure à 20 ha et inférieure à 25 ha

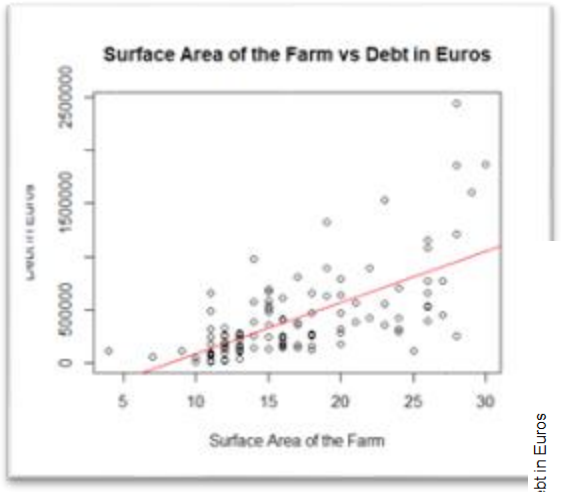
- Dependent variable (y variables) = amount of debt farm has in euros
  - Testing for normality of the dependent variable

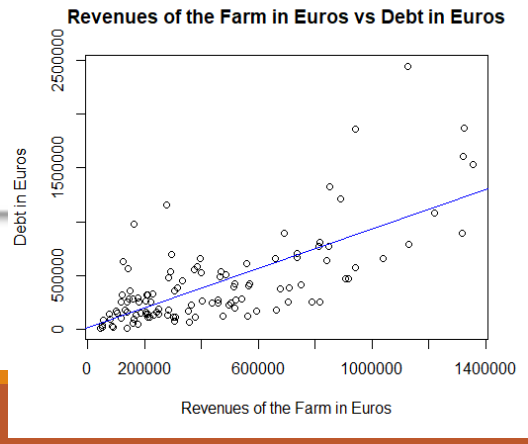
```
        Shapiro-Wilk normality test

data:  data$Debt.in.EurosSW
W = 0.98795, p-value = 0.4078
```

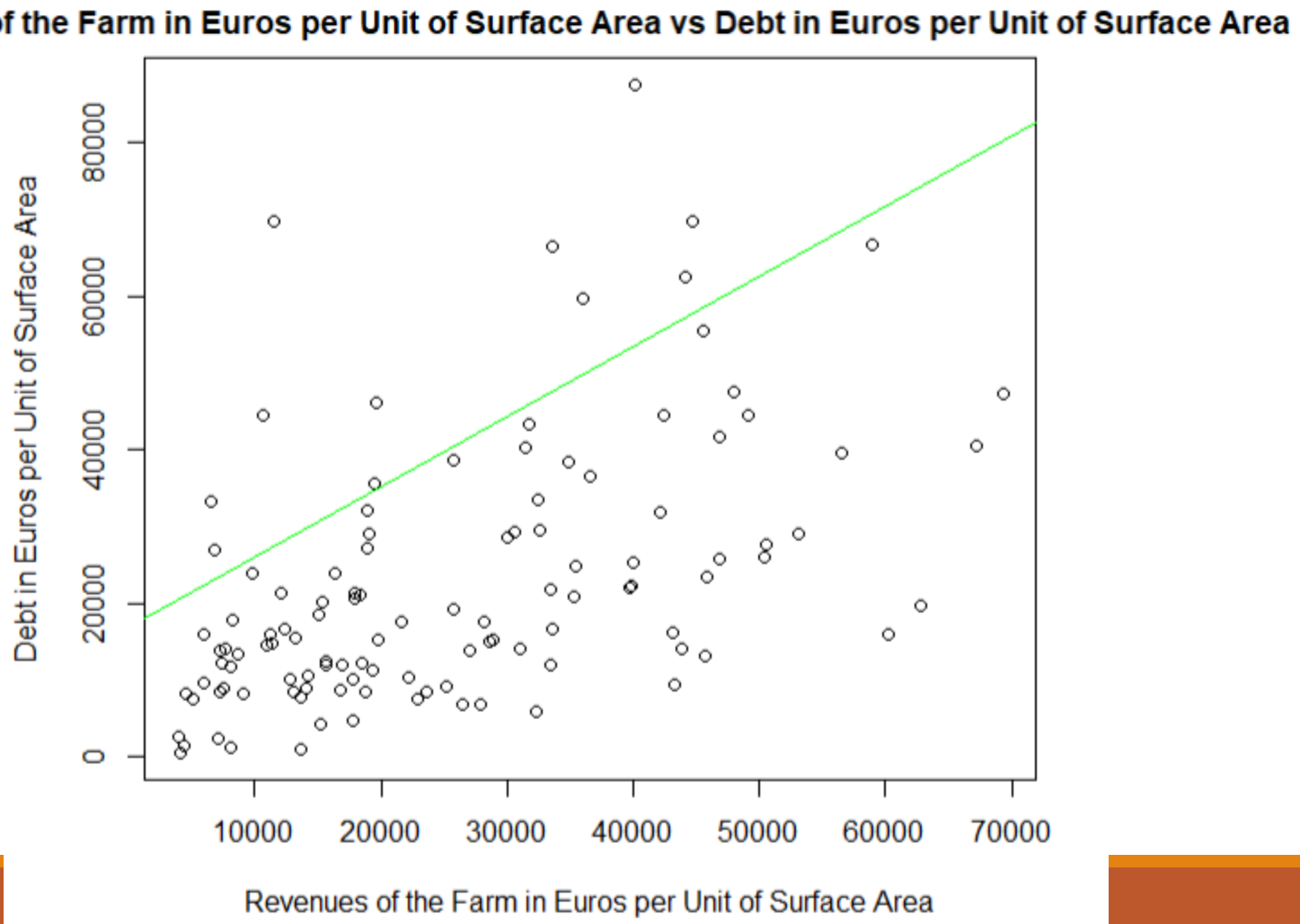
P value is **greater** than 0.05 therefore considered normal.

1) Does the farm's size (in terms of surface area and revenue) impact the quantity of debt the "**Mixed farming operations with granivores**" farms have?

```
> plot(data$Surface.area.of.the.Farm, data$Debt.in.Euros,
+      xlab = "Surface Area of the Farm", ylab = "Debt in Euros",
+      main = "Surface Area of the Farm vs Debt in Euros")
> abline(model, col = "red")
```







1) Does the farm's size (in terms of surface area and revenue) impact the quantity of debt the "**Mixed farming operations with granivores**" farms have?

```
> model <- lm(Debt.in.Euros...SA ~ Revenues.SA, data = data)
> summary(model)

Call:
lm(formula = Debt.in.Euros...SA ~ Revenues.SA, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-25493  -9154  -3859   5556  56949

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.813e+03  2.626e+03   3.356  0.00108 **
Revenues.SA 5.406e-01  8.693e-02   6.219 8.85e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14790 on 112 degrees of freedom
Multiple R-squared:  0.2567,    Adjusted R-squared:  0.2501
F-statistic: 38.68 on 1 and 112 DF,  p-value: 8.847e-09
```

Multiple R-Squared= 0.2567, indicating that about 25.67% of the variability in the dependent variable is explained by the independent variable

F-statistic= 38.68 (1 and 112 degrees of freedom. & associated p-value is quite low and less than o.o5 (8.847e-09).
This indicates that the model is statistically significant.

2) Does the sex of the farmer impact the quantity of debt the farmer has?

- Chosen test: Welch two sample t-test
- Explanatory variables (ID_sex) = gender of farm owner (category)

2297	SEXEP	blanc	Sans objet
2298	SEXEP	1	Homme
2299	SEXEP	2	Femme

- Dependent variable (ID_debt) = amount of debt farm has in euros

```
> data <- read_csv("project_data.csv")
Rows: 244 Columns: 2
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
dbl (2): ID_sex, ID_debt

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
>
> data <- data %>%
+     mutate(ID_sex = factor(ID_sex, levels = c(1, 2), labels = c("Male", "Female")),
+            ID_debt = as.numeric(ID_debt))
> t_test_result <- t.test(ID_debt ~ ID_sex, data = data)
> print(t_test_result)

        Welch Two Sample t-test

data:  ID_debt by ID_sex
t = 0.95524, df = 28.042, p-value = 0.3476
alternative hypothesis: true difference in means between group Male and group Female is not equal to 0
95 percent confidence interval:
 -49544.64 136142.57
sample estimates:
  mean in group Male mean in group Female
            308049.6             264750.7
```

- Result: The t-test shows no statistically significant difference in debt levels between male and female farmers. High variability in the data and the p-value suggest any observed differences might be due to chance, not gender disparity.

Visualization of result (2)

Code used for the Boxplot:

```
> library(ggplot2)
> ggplot(data, aes(x = ID_sex, y = ID_debt, fill = ID_sex)) +
+     geom_boxplot() +
+     labs(title = "Boxplot of Debts by Gender",
+          x = "Gender",
+          y = "Debt (Euros)") +
+     theme_minimal()
```

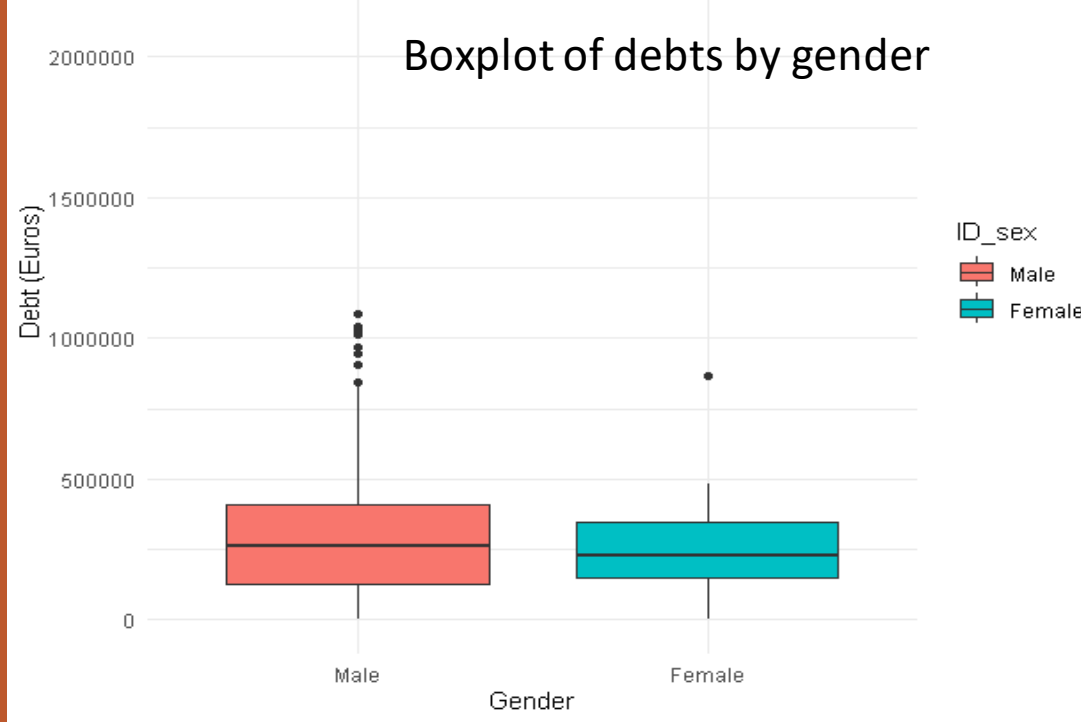
Code used for the Density plot:

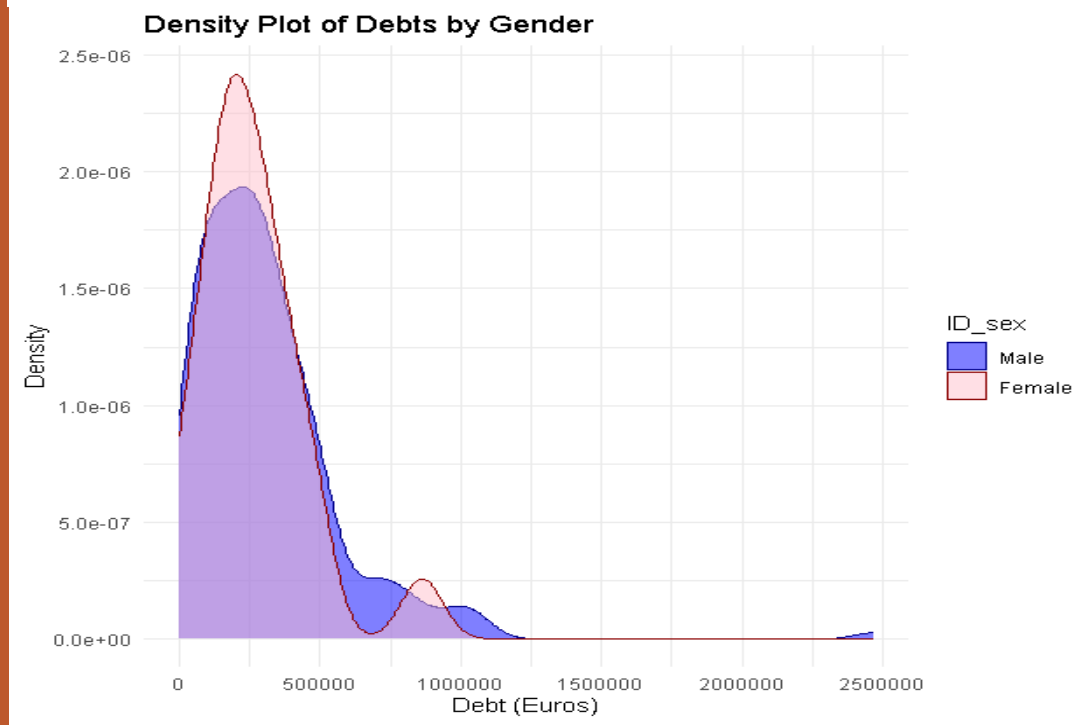
```
> ggplot(data, aes(x = ID_debt, fill = ID_sex, color = ID_sex)) +
+     geom_density(alpha = 0.5) +
+     labs(title = "Density Plot of Debts by Gender",
+          x = "Debt (Euros)",
+          y = "Density") +
+     scale_fill_manual(values = c("blue", "pink")) +
+     scale_color_manual(values = c("darkblue", "darkred")) +
+     theme_minimal()
```

Male – female persentage count:

```
> data <- read.csv("project_data.csv")
> sex_counts <- table(data$ID_sex)
>
> sex_proportions <- prop.table(sex_counts)
> sex_percentages <- sex_proportions * 100
> print(sex_percentages)

       1        2
91.393443  8.606557
> names(sex_percentages) <- c("Male", "Female")
> print(sex_percentages)
     Male    Female
91.393443  8.606557
```

We also need to metion that in selected production type farms there are only 8,6% of females. And again - high variability in the data and the p-value suggest any observed differences might be due to chance, not gender disparity.





3) Does the age of the farmer impact the quantity of debt the farmer has ?

Hypothesis :

1/ The oldest farmer managers may have a better management with the debt thanks to them seniority

2/ The youngest may have a better management due to the school they were in, or because they have more access to information with Internet, etc.

3) Does the age of the farmer impact the quantity of debt the farmer has ?

Now, we want to determine if there is a relationship between the debts dans the age range of the farm manager. First we do a correlation test.

For that, we had to import the two parameters from the database.

```{r}
cor.test(as.numeric(data_age_debt...Sheet1$ID_age),data_age_debt...Sheet1$ID_debt)
```

```
        Pearson's product-moment correlation

data:  as.numeric(data_age_debt...Sheet1$ID_age) and
data_age_debt...Sheet1$ID_debt
t = 0.28067, df = 242, p-value = 0.7792
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1077908  0.1433004
sample estimates:
       cor
0.01803924
```

After this correlation test, we can see that the value of the corr is near of 0 than 1 or -1. Both have not got any relationship.

3) Does the age of the farmer impact the quantity of debt the farmer has ?

To verify this result, we can do a chi-square test.

We pose H0 : Variables are independents

and H1 : Variables are dependents

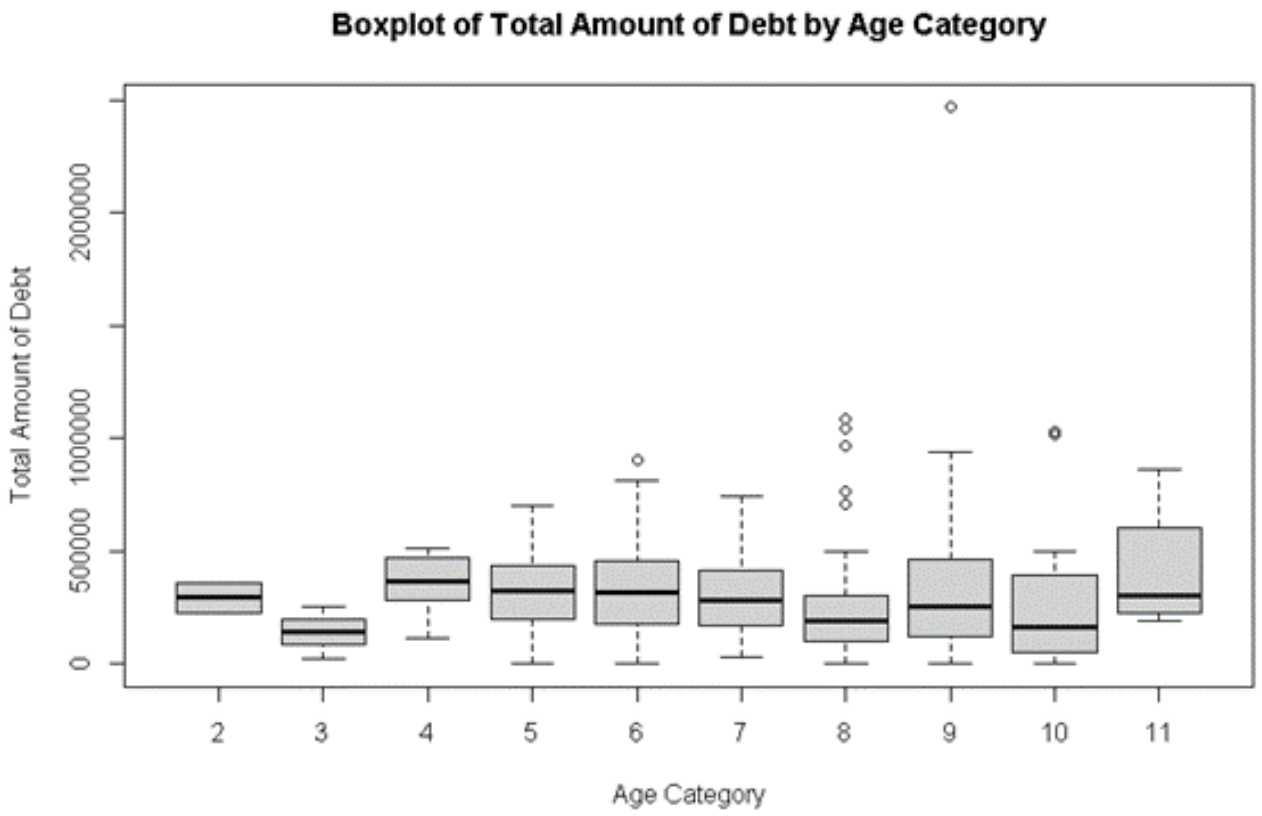
If pvalue < 0,05 we reject H0

```{r}
chisq.test(x = data_age_debt...Sheet1$ID_age,y= data_age_debt...Sheet1$ID_debt)
```

```
Avis dans chisq.test(x = data_age_debt...Sheet1$ID_age, y =
data_age_debt...Sheet1$ID_debt) :
  L'approximation du Chi-2 est peut-être incorrecte

        Pearson's Chi-squared test

data:  data_age_debt...Sheet1$ID_age and data_age_debt...Sheet1$ID_debt
X-squared = 2196, df = 2187, p-value = 0.442
```

pvalue > 0,05 we keep H0, both variables are independents.

Visualization with a Boxplot :



3) Does the age of the farmer impact the quantity of debt the farmer has ?

HERE, YOU CAN SEE A BOXPLOT SHOWING THE DIFFERENT PARAMETERS OF EACH CATEGORY.

AS WE CAN SEE, THE VARIOUS BOXPLOTS ARE ALL DIFFERENT FROM ONE ANOTHER: SOME FOLLOW A NORMAL DISTRIBUTION, OTHERS NOT AT ALL; SOME HAVE EXTREME VALUES, WHILE OTHERS DO NOT. AND SOME HAVE OUTLIERS.

TO CONCLUDE THIS QUESTION, THERE IS NO RELATIONSHIP BETWEEN THE AGE RANGE OF THE FARM MANAGER AND THE QUANTITY OF DEBT OF THE FARM.

4) Do farms present on natura 2000 zones have more debt? # qualitativeanalysis

- t-test goal : compare the means of farms that have or do not have an environmental constraint (represented by 1 and 0 and in the ZENVI column) with the mean of the debt column called TETD3. If p-value > 0.05, our results are insignificant.

```
        Welch Two Sample t-test

data:  RICA_Data_CSV$ZENVI and RICA_Data_CSV$TDTE3
t = -60.573, df = 7202, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -250792.0 -235068.4
sample estimates:
   mean of x    mean of y
3.984451e-02 2.429302e+05
```

ZENVI	0	LamajeurepartiedelaSAUn'estpasenzoneNatura2000
ZENVI	1	LamajeurepartiedelaSAUestsituéeenzoneNatura2000

- Chi-square test: test the independence of two samples. If the p-value > or = 0.05 we accept the HO hypothesis.

```
        Pearson's Chi-squared test

data:  RICA_Data_CSV$ZENVI and RICA_Data_CSV$TDTE3
X-squared = 7152, df = 7158, p-value = 0.5179
```

*Zone 2000 - A development project or specific area in France established or recognized around the year 2000.

4) Do farms present on natura 2000 zones have more debt?

- Chosen test: Wilcoxon rank sum test
- Explanatory variables (ID_zone) = belonging of the farm to *zone 2000 (category)

ZENVI	0	LamajeurepartiedelaSAUn'estpasenzoneNatura2000
ZENVI	1	LamajeurepartiedelaSAUestsituéeenzoneNatura2000

- Dependent variable (ID_debt) = amount of debt farm has in euros

```
> data <- read.csv("zone_data.csv")
> zone_2000 <- data[data$ID_zone == 1, "ID_debt"]
> not_in_zone_2000 <- data[data$ID_zone == 0, "ID_debt"]
> wilcox.test(ID_debt ~ ID_zone, data = data)

        Wilcoxon rank sum test with continuity correction

data:  ID_debt by ID_zone
W = 3009, p-value = 0.07281
alternative hypothesis: true location shift is not equal to 0
```

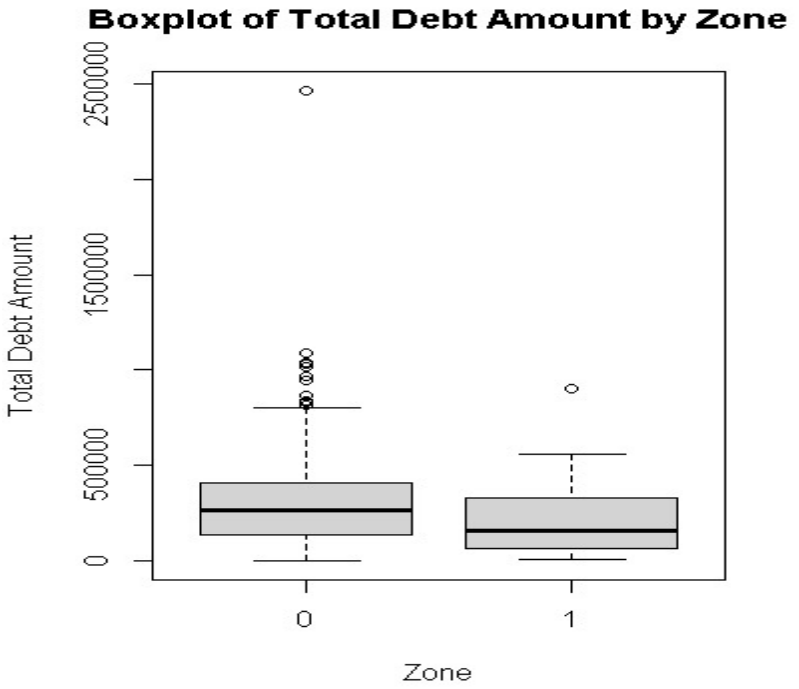
- Result: based on the Wilcoxon rank sum test, there is no significant difference in the median total debt amounts between farms in zone 2000 and farms not in zone 2000.

*Zone 2000 - A development project or specific area in France established or recognized around the year 2000.

Visualization of result (4)

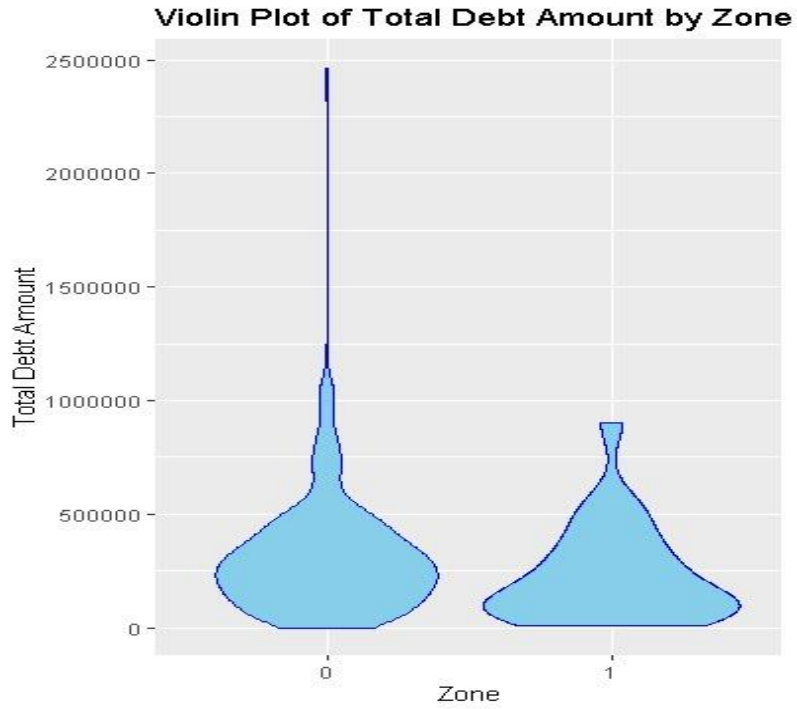
Code used for the Boxplot:

```
> boxplot(ID_debt ~ ID_zone, data = data,
+         xlab = "Zone", ylab = "Total Debt Amount",
+         main = "Boxplot of Total Debt Amount by Zone")
```



Code used for the Violin plot:

```
> library(ggplot2)
> ggplot(data, aes(x = factor(ID_zone), y = ID_debt)) +
+     geom_violin(fill = "skyblue", color = "blue") +
+     labs(x = "Zone", y = "Total Debt Amount",
+          title = "Violin Plot of Total Debt Amount by Zone")
```



People have more debts in zones that have no environmental constraints than in those who do have some, but the difference is not really significant. Thus, there's no correlation between Natura 2000 zones and debts.

Conclusions

- Larger farms generally have more debt than smaller farms ;
  - More borrowed investment for expansion ie; new technologies, more land, modern equipment etc…
  - Risky sector due to market fluctuations, weather events, diseases…

(USDA, n.d.)

- Men are more frequently in debt than women. However, this can be explained by the fact that there are more farms owned by men in France; only 31. 6% of farmers are women in France. (Eurostat, n.d.)

- Farms who are not on Natura 2000 zones are faced with more debts but the difference is not significance. However, once again, there are a lot less farms on Natura 2000 zones compared to 'normal' zones hence, this statement is not conclusive.

Limitations

- Only one production type was observed
- Gender and age for qualitative data since other factors impact it (social factors etc...)
- Only certain characteristics were looked at
- More multivariate analysis could be carried out

Sources

Viande.Info. (n.d.). *Situation financière et santé des éleveurs*. Accueil. https://www.viande.info/eleveur-situation-financiere-sante-endettement-suicide

Larger farms and younger farmers are more vulnerable to financial stress. USDA ERS - Larger Farms and Younger Farmers Are More Vulnerable to Financial Stress. (n.d.). https://www.ers.usda.gov/amber-waves/2019/october/larger-farms-and-younger-farmers-are-more-vulnerable-to-financial-stress/

Europa. Statistics Explained. (n.d.). https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=431368