



Mémoire : SAE PROJET

2023

**Prêts garantis par
l'État"**

Thomas Belaid
Sami Many Kara

Melissa Djenadi
Bastien Ebely





Mémoire : SAE PROJET

2023

**Prêts garantis par
l'État"**

Thomas Belaid

XXXXX XXXXX



XXXXXX XXXX

XXXXXX XXXX

SOMMAIRE

- INTRODUCTION DU SUJET
- INTRODUCTION DE LA SAE
- PRÉSENTATION DE LA BASE DONNÉES
- DESCRIPTIONS DES DONNÉES BRUTES
- CRÉATION ET NETTOYAGE DE LA BASE DONNÉE
- ANALYSE STATISTIQUES UNIVARIÉES
- ANALYSES STATISTIQUES BIVARIÉES
- RÉGRESSION LINÉAIRE
- CONCLUSION SCIENTIFIQUE

INTRODUCTION DU SUJET

La pandémie de COVID-19, provoquée par le virus SARS-CoV-2, a profondément marqué le monde entier depuis son émergence à la fin de l'année 2019. Ce nouveau coronavirus a rapidement traversé les frontières, se propageant à une vitesse alarmante et entraînant des conséquences sanitaires, économiques et sociales sans précédent. Avec des millions de personnes infectées et des pertes tragiques en vies humaines, la pandémie a mis en lumière les fragilités de nos sociétés et a imposé des défis inédits aux gouvernements, aux systèmes de santé et aux économies mondiales.

La pandémie de COVID-19 a entraîné des répercussions majeures sur tous les aspects de la vie, notamment sur la santé publique, l'économie mondiale et les dynamiques sociales. En France, les conséquences économiques de la pandémie ont été significatives, avec des entreprises de tous les secteurs luttant pour survivre face aux restrictions imposées et à la baisse de la demande. Face à cette situation critique, le gouvernement français a mis en place un ensemble de mesures de soutien pour atténuer les effets néfastes de la crise.

La France, tout comme de nombreux autres pays, a été fortement touchée par les retombées économiques de la pandémie. Face à l'ampleur des difficultés rencontrées par les entreprises françaises.

Parmi ces mesures, les prêts garantis par l'État ont été une réponse clé pour aider les entreprises françaises à faire face aux difficultés financières causées par la pandémie. Ces prêts ont été conçus pour offrir une certaine stabilité et une sécurité financière aux entreprises en difficulté. L'idée centrale derrière les prêts garantis par l'État était de fournir des liquidités aux entreprises afin de les aider à maintenir leurs activités, à payer leurs employés et à répondre à leurs obligations financières, même en période de crise économique.

Ces prêts ont été distribués à travers toute la France, mais leur répartition n'a pas été uniforme. Différents facteurs géographiques et économiques ont influencé la manière dont les prêts ont été accordés et répartis entre les entreprises. Par exemple, certaines régions ont peut-être été plus durement touchées par la pandémie, ce qui a pu se refléter dans une plus grande demande de prêts garantis par l'État. De même, certains secteurs d'activité ont été plus gravement impactés par les restrictions et les mesures de confinement, ce qui a pu entraîner une concentration plus élevée de prêts dans ces domaines soutenir les entreprises en difficulté et stimuler la reprise économique à long terme.

INTRODUCTION DE LA SAE

Dans le cadre de cette situation d'apprentissage (SAE), l'objectif principal est d'analyser la répartition des prêts garantis par le gouvernement français aux entreprises touchées par les conséquences économiques de la pandémie. Pour mener à bien cette étude, nous avons sélectionné un ensemble de données brutes qui comprend plusieurs variables liées à ces prêts.

Les prêts garantis par le gouvernement français ont été mis en place dans le but d'aider les entreprises françaises à faire face aux difficultés économiques causées par la pandémie. Le concept de ces prêts repose sur la fourniture de prêts avec des garanties de l'État, offrant ainsi une certaine sécurité aux entreprises. Le montant des prêts accordés dépend de divers critères, tels que la localisation géographique de l'entreprise, le secteur d'activité et le domaine d'activité.

Nous avons particulièrement utilisé la variable appelée "montant total" pour mettre en évidence les variations de ce montant en fonction des différentes variables. Cette variable nous permet d'analyser le total des prêts accordés aux entreprises en fonction de leur situation géographique et de leur domaine d'activité. En l'étudiant, nous espérons identifier des tendances et tirer des conclusions sur la répartition des prêts garantis par l'État en France.

En examinant les données disponibles, nous souhaitons comprendre comment ces prêts sont répartis à travers le pays et comment cette répartition a été influencée par divers facteurs géographiques et économiques. Par exemple, nous pourrions examiner si les entreprises situées dans certaines régions ont reçu une plus grande part de prêts par rapport à d'autres régions, ou si des secteurs d'activité spécifiques ont été davantage soutenus par ces prêts. Cette analyse nous permettra d'obtenir une meilleure compréhension de l'impact économique de la pandémie en France.

En résumé, cette SAE vise à étudier la répartition des prêts garantis par l'État en France et à comprendre comment cette répartition a été influencée par des facteurs géographiques et économiques. En analysant les données sur les prêts accordés aux entreprises, nous espérons identifier des tendances et des conclusions significatives qui contribueront à une meilleure compréhension de l'impact de la pandémie sur l'économie française.

PRÉSENTATION DE LA BASE DE DONNÉES :

Nous disposons d'un jeu de données exhaustif traitant des prêts accordés par l'État pendant la pandémie de COVID-19 jusqu'au 31 décembre 2021, ce qui en fait un rapport annuel complet. Ce jeu de données est composé de dix colonnes qui représentent différentes variables essentielles pour notre étude approfondie.

La première variable, "*dispositif*", désigne l'objet de l'étude que nous allons analyser en détail. La deuxième variable, "*nombre_page*", représente le nombre de prêts accordés dans le cadre du dispositif spécifié. La troisième variable, "*montant_total*", indique le montant total des prêts accordés.

Les variables "*reg*" et "*libelle_region*" fournissent des informations sur la région associée à chaque prêt. De même, les variables "*dep*" et "*libelle_département*" décrivent le département correspondant à chaque prêt.

La variable "*code_section*" est liée à la classification des prêts selon une certaine section, tandis que la variable "*libelle_section*" fournit une description plus détaillée de la section en question.

Enfin, la variable "*date_maj*" indique la date de mise à jour des informations relatives à chaque prêt dans la base de données.

Dans l'ensemble, nous disposons de 1747 observations, ce qui représente une quantité substantielle de données pour mener notre analyse statistique et obtenir des insights précieux sur les prêts accordés pendant cette période cruciale de la pandémie.

DESCRIPTIONS DES DONNÉES BRUTES :

Dispositif : C'est une variable qualitative nominale qui indique le type de dispositif associé aux données, qui est toujours le même, à savoir "Prêts garantis par l'État". Il s'agit donc d'une variable catégorique qui décrit la nature du programme ou du système dont les données sont extraites.

Nombre_page : C'est une variable quantitative continue. Elle représente une mesure numérique qui peut prendre des valeurs décimales et peut varier sur un continuum. Elle indique le nombre de PGE (Prêts Garantis par l'État) associés à chaque entrée de données.

Montant_total : C'est une variable quantitative continue. Elle représente une mesure numérique qui peut prendre des valeurs décimales et peut varier sur un continuum. Cette variable indique le montant total associé à une certaine mesure ou caractéristique dans le jeu de données, tel que le montant total des ventes réalisées ou le budget total alloué.

Reg : C'est une variable quantitative discrète. Elle représente un code ou un identifiant numérique pour une région spécifique. Les valeurs de cette variable incluent tous les numéros de régions présents dans le jeu de données. Elle est utilisée pour identifier la région à laquelle chaque entrée de données est associée.

Libelle_région : C'est une variable qualitative nominale. Elle représente le nom ou la description de la région correspondant au code de région (variable "Reg"). Les valeurs de cette variable incluent tous les noms des régions présentes dans le jeu de données. Elle fournit des informations textuelles sur la région associée à chaque entrée de données.

Dep : C'est une variable quantitative discrète. Elle représente un code ou un identifiant numérique pour un département spécifique. Les valeurs de cette variable incluent tous les numéros de départements présents dans le jeu de données. Elle est utilisée pour identifier le département auquel chaque entrée de données est associée.

Libelle_département : C'est une variable qualitative nominale. Elle représente le nom ou la description du département correspondant au code de département (variable "Dep"). Les valeurs de cette variable incluent tous les noms des départements présents dans le jeu de données. Elle fournit des informations textuelles sur le département associé à chaque entrée de données.

Code_section : C'est une variable qualitative nominale. Elle représente un code ou un identifiant pour une section spécifique dans les secteurs d'activité. Les valeurs de cette variable incluent toutes les sections des secteurs d'activité présentes dans le jeu de données. Elle peut être utilisée pour classer ou regrouper les entrées de données en fonction d'une certaine section ou catégorie.

Libelle_section : C'est une variable qualitative nominale. Elle représente le nom ou la description de la section correspondant au code de section (variable "Code_section"). Les valeurs de cette variable incluent toutes les sections des secteurs d'activité présentes dans le jeu de données. Elle fournit des informations textuelles sur la section associée à chaque entrée de données.

CRÉATION DE LA BASE DE DONNÉES :

Nous avons remarqué que la base de données n'est pas nettoyée et ordonnée, nous avons donc effectué certaines modifications afin d'avoir une meilleure base de données et que l'on puisse travailler plus efficacement.

```
#Changer le nom des colonnes
```

```
names(Base_de_Donnée)[names(Base_de_Donnée) == "dispositif"] <- "Dispositif"
names(Base_de_Donnée)[names(Base_de_Donnée) == "nombre_pge"] <- "Nombre_de_Page"
names(Base_de_Donnée)[names(Base_de_Donnée) == "montant_total"] <- "Montant_Total"
names(Base_de_Donnée)[names(Base_de_Donnée) == "reg"] <- "Region"
names(Base_de_Donnée)[names(Base_de_Donnée) == "libelle_region"] <- "Nom_Region"
names(Base_de_Donnée)[names(Base_de_Donnée) == "dep"] <- "Département"
names(Base_de_Donnée)[names(Base_de_Donnée) == "libelle_departement"] <- "Nom_Département"
names(Base_de_Donnée)[names(Base_de_Donnée) == "code_section"] <- "Code_Section"
names(Base_de_Donnée)[names(Base_de_Donnée) == "libelle_section"] <- "Nom_Section"
```

Pour cela, nous avons changé les titres des colonnes afin d'avoir des titres plus précis et plus compréhensibles.

Nous avons utilisé `names(Data)`
`[names(Data) == Nom de la colonne] = Nom de la colonne`

	A	B	C	D	E	F	G	H	I	J
1	dispositif	nombre_pge	montant_total	reg	libelle_region	dep	libelle_departement	code_section	libelle_section	date_maj



Dispositif	Nombre_de_Page	Montant_Total	Region	Nom_Region	Département	Nom_Département	Code_Section	Nom_Section	date_maj
------------	----------------	---------------	--------	------------	-------------	-----------------	--------------	-------------	----------

Afin de pouvoir manipuler certaines données, nous avons dû modifier le format de certaines données. Exemple, nous avons modifié, le format de la variable montant total au format monétaire. Concernant la variable "Montant totale", nous avons modifié le format de cette variable afin de pouvoir faire des manipulations statistiques (test, Statistique univariée, Statistique bivariée...) pour cela, nous avons utilisé une librairie, la librairie stringr et on a utilisé la fonction : `As.numeric =`
Alors nous avons remplacé les "," par des ".". Afin que les « caractères » se transforment en « numérique ». Nous avons remarqué qu'une colonne ne servait pas beaucoup à notre base de données, c'est celle qu'on a renommé « Date de dernière mise à jour », car elle indique la même valeur, c'est-à-dire « 31-12-2021 ». Pour cela, nous avons utilisé une fonction de R. La fonction `subset()`.

```
# Enlever la colonne inutile
```

```
Base de Donnée <- subset(Base de Donnée, select = -c(Base de Donnée$date_maj))
```

ANALYSE STATISTIQUE :

L'analyse statistique consiste à collecter et à examiner tous les échantillons de données extraits d'un ensemble de données. En statistique, un échantillon est une sélection représentative d'une population.

ANALYSES STATISTIQUES UNIVARIÉES :

L'analyse statistique univariée se concentre sur l'étude d'une seule variable à la fois dans un ensemble de données. Elle vise à explorer et à décrire les caractéristiques de cette variable en utilisant des mesures statistiques telles que la moyenne, la médiane, la variance, l'écart-type, etc. L'analyse statistique univariée permet de comprendre la distribution, la tendance centrale, la dispersion et la forme de la variable étudiée. Elle fournit des informations précieuses sur les propriétés intrinsèques de la variable, ce qui permet d'identifier les valeurs atypiques, les schémas ou les tendances spécifiques qui peuvent être présents. Cela permet également de faire des comparaisons entre différentes populations ou groupes sur la base de cette variable, en utilisant des tests statistiques appropriés. En résumé, l'analyse statistique univariée offre une vision approfondie d'une seule variable, ce qui permet de mieux comprendre ses caractéristiques et son rôle dans l'ensemble des données.

Variable quantitative :

Lorsque nous nous confrontons à des variables quantitatives, qui se réfèrent à des mesures numériques effectuées sur une échelle continue, il est possible de recourir à la fonction `summary()` en vue d'obtenir des mesures descriptives essentielles. Cette fonction est fréquemment employée dans le domaine de l'analyse statistique en vue de délivrer une synthèse concise, mais pertinente, des caractéristiques inhérentes aux variables quantitatives.

VARIABLES QUANTITATIVES :

POUR LA VARIABLE NOMBRE DE PRÊT PAR L'ÉTAT :

Nous avons utilisé la fonction de R `summary()`,

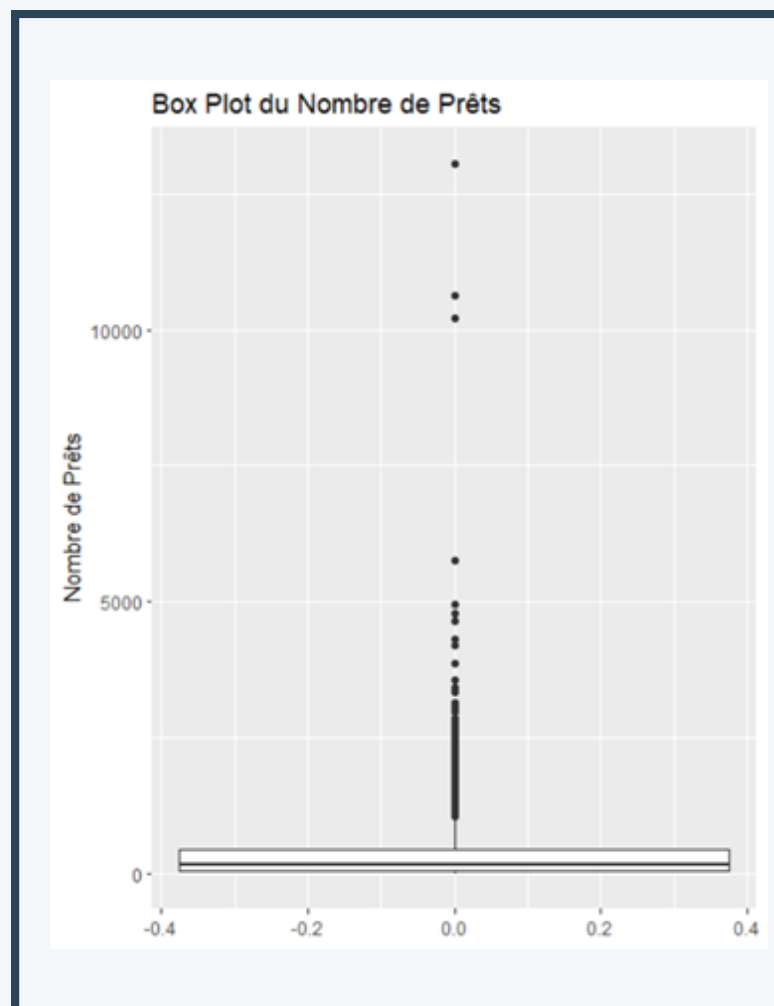
Les résultats sont :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.0	46.0	152.0	400.3	449.0	13060.0

- Nous pouvons voir que le nombre de Prêt par l'état minimum est 3
- Le nombre de Prêt par l'état maximum est 13060
- La moyenne du nombre de Prêt par l'état est 400.3
- La Médiane du nombre de Prêt par l'état est 152.2

Nous avons réalisé une boîte à moustache afin d'observer la dispersion de cette variable

Nous pouvons voir que la boîte à moustache est illisible, mais surtout qu'il y a beaucoup de point en dehors des moustaches de la boîte, ce sont des valeurs statistiquement aberrantes ce qui montre qu'il y a une certaine inégalité de la répartition des prêts.



VARIABLES QUANTITATIVES :

POUR LA VARIABLE MONTANT TOTAL :

Nous avons utilisé la fonction de R `summary()`,

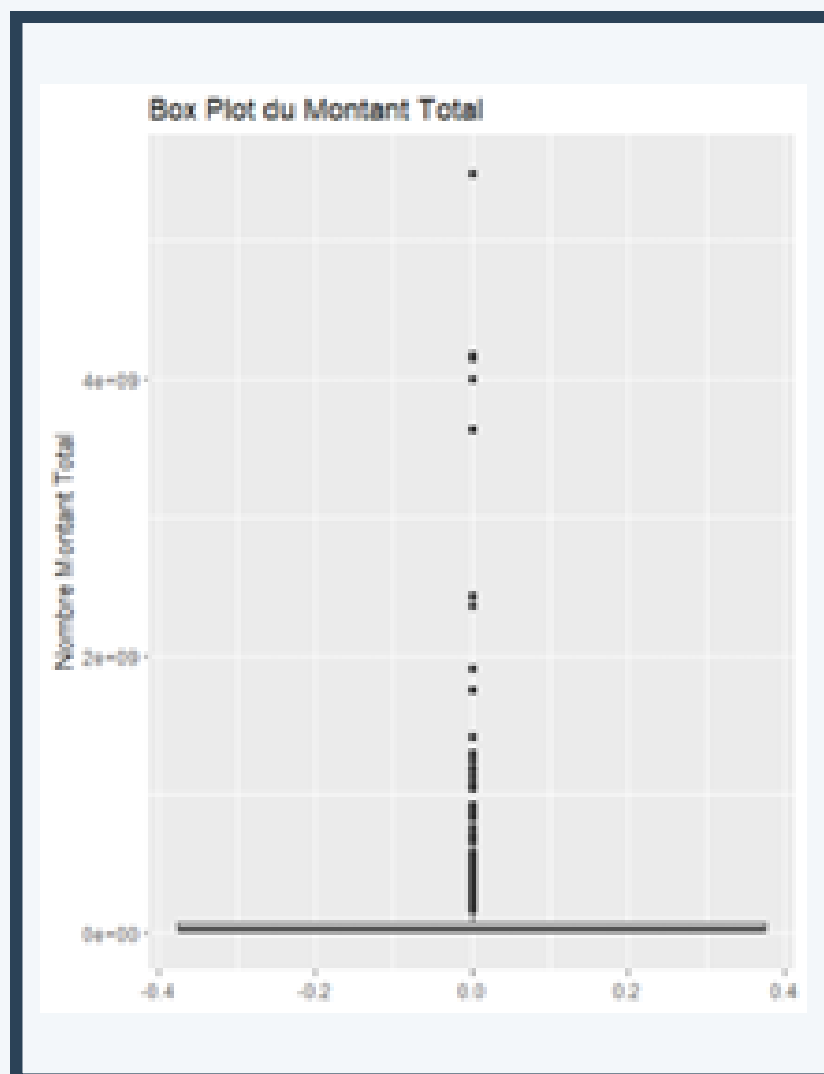
Les résultats sont :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.800e+04	4.532e+06	1.666e+07	8.188e+07	6.576e+07	5.487e+09

- Le montant total minimum est $2.800 \cdot 10^4$
- Le montant total maximum est $5.487 \cdot 10^9$
- La moyenne du montant total est $8.188 \cdot 10^7$
- La médiane du montant total est $1.666 \cdot 10^7$

Nous avons réalisé une boîte à moustache afin d'observer la dispersion de cette variable

Nous pouvons voir que la boîte à moustache est illisible mais surtout qu'il y a beaucoup de point en dehors des moustaches de la boîte, ce sont des valeurs statistiquement aberrantes ce qui montre qu'il y a une certaine inégalité de la répartition des montants totaux.



VARIABLES QUALITATIVES :

Pour ces variables, nous avons sélectionné certaines variables, car les variables qualitatives vont de pair afin de ne pas avoir une répartition, nous allons étudier les libellés des variables :

- Région
- Département
- Section

Pour ces variables, nous allons utiliser plusieurs fonctionnalités et fonction de R afin d'avoir une étude précise de celle-ci.

- La fonction `table()` en R permet de créer un tableau de fréquences qui compte le nombre d'occurrences de chaque valeur d'une variable.
- La library `ggplot2` qui nous permet de faire des graphiques de qualités.
- La fonction `sort()` qui peut nous permettre de mettre les valeurs dans l'ordre décroissant

Nous allons montrer la méthode pour une variable afin de ne pas avoir de répétition.

Pour la Variable Nom de Région :

Nous avons créé une table qui contient les effectifs de chaque région qui apparaissent dans la base de données. La fonction `table()` nous a permis d'avoir les effectifs et les avons implémenter dans une base de données appelées `Information_Région` .

```
#Création de la table contenant les effectifs |  
Information_Région = table(Base_de_Donnée$Nom_Region)
```

Nous avons trié la table en fonction de l'effectif dans l'ordre décroissant notamment avec la fonction `sort()`.

```
# Mettre dans l'ordre décroissant en fonction de l'effectif  
Information_Région <- sort(Information_Région, decreasing = TRUE)  
View(Information_)
```

VARIABLES QUALITATIVES :

Pour la variable Région :

Nous avons créé la table Information_Région avec la même méthode que celle précédente.

Nous avons comme table :

	Var1	Freq
1	Occitanie	225
2	Auvergne-Rhône-Alpes	211
3	Nouvelle-Aquitaine	208
4	Grand Est	171
5	Île-de-France	139
6	Bourgogne-Franche-Comté	136
7	Provence-Alpes-Côte d'Azur	105
8	Centre-Val de Loire	101
9	Hauts-de-France	86
10	Pays de la Loire	85
11	Normandie	84
12	Bretagne	68
13	Corse	34
14	La Réunion	19
15	Guadeloupe	18
16	Guyane	18
17	Martinique	17

ring 1 to 17 of 18 entries, 2 total columns

Afin d'améliorer la visualisation des données, nous avons créé un histogramme qui représente l'effectif de chaque région. Pour éviter d'avoir un histogramme illisible avec trop de barres, nous avons choisi de limiter notre analyse aux 10 régions les plus représentées. Cette approche nous permet de focaliser notre attention sur les régions les plus importantes en termes de fréquence et d'effectif, ce qui facilite l'interprétation et la compréhension des résultats. En se concentrant sur les régions les plus significatives, nous sommes en mesure d'identifier plus facilement les tendances et les disparités entre les différentes régions, ce qui peut être utile dans une analyse plus approfondie des données.

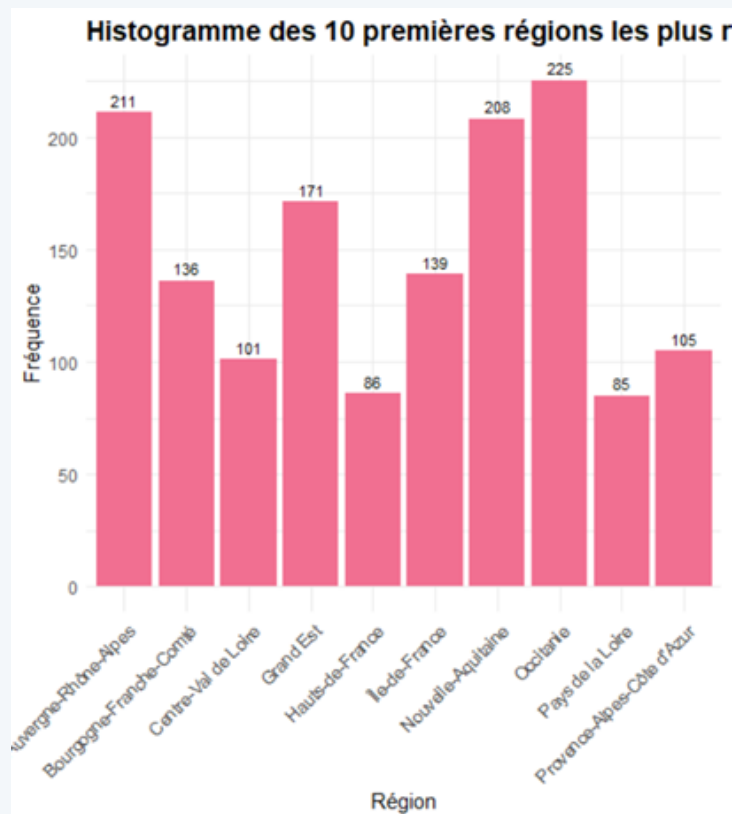
```
# Limité à 10 pays
df <- data.frame(Région = names(Information_Région)[1:10],
                  Fréquence = as.numeric(Information_Région)[1:10])
```

Afin d'obtenir une meilleure qualité pour notre histogramme, nous avons utilisé la librairie ggplot2. Cette librairie offre des fonctionnalités avancées pour la création de graphiques, ce qui nous permet de personnaliser facilement l'apparence de notre histogramme.

```
# Création de l'histogramme
ggplot(df, aes(x = Région, y = Fréquence)) +
  geom_bar(stat = "identity", fill = "lightblue") +
  labs(title = "Histogramme des 10 premières region les plus représenté",
       x = "Région",
       y = "Fréquence") +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, face = "bold"),
        axis.text.x = element_text(angle = 45, hjust = 1))
```


VARIABLES QUALITATIVES :

Histogramme :



L'analyse du graphique révèle que la région la plus représentée est la région d'Occitanie, suivie de près par la région Auvergne-Rhône-Alpes et la région Nouvelle-Aquitaine. Ces régions se distinguent par leur fréquence élevée dans notre ensemble de données. La région d'Occitanie affiche la fréquence la plus élevée, ce qui indique qu'elle est largement représentée dans nos observations.

VARIABLES QUALITATIVES :

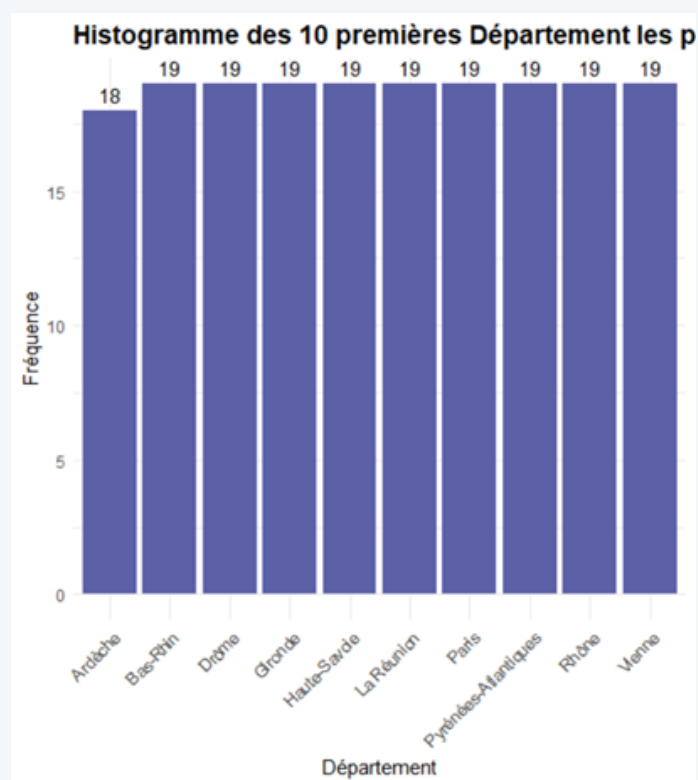
Pour la variable Nom Département :

Nous avons créé la table Information_Département avec la même méthode que celle précédente.

▲	Var1	Freq
1	Bas-Rhin	19
2	Drôme	19
3	Gironde	19
4	Haute-Savoie	19
5	La Réunion	19
6	Paris	19
7	Pyrénées-Atlantiques	19
8	Rhône	19
9	Vienne	19
10	Ardèche	18
11	Aude	18
12	Bouches-du-Rhône	18
13	Côte-d'Or	18
14	Dordogne	18
15	Gard	18
16	Guadeloupe	18
17	Guyane	18

Nous pouvons voir qu'il n'y a pas une grande dispersion entre les Départements. La faible dispersion entre les départements est confirmée par l'analyse de l'histogramme des données, qui montre une distribution relativement uniforme des fréquences.

Histogramme :



Les départements les plus représentés, avec un effectif de 19 chacun, sont : Bas-Rhin, Drôme, Gironde, Haute-Savoie, La Réunion, Paris, Pyrénées-Atlantiques, Rhône et Vienne. Ces départements se distinguent par leur fréquence élevée dans notre ensemble de données, ce qui indique leur importance dans notre analyse.

VARIABLES QUALITATIVES :

Pour la variable Nom de Section :

Nous avons créé la table `Information_Section` avec la même méthode que celle précédente.

	Var1	Freq
1	Agriculture, sylviculture et pêche	102
2	Autres activités de services	102
3	Commerce	102
4	Construction	102
5	Hébergement et restauration	102
6	Industrie manufacturière	102
7	Activités de services administratifs et de soutien	101
8	Activités spécialisées, scientifiques et techniques	101
9	Arts, spectacles et activités récréatives	101
10	Enseignement	101
11	Santé humaine et action sociale	101
12	Transports et entreposage	101
13	Activités financières et d'assurance	100
14	Activités immobilières	100
15	Information et communication	99
16	Non connu	98
17	Production et distribution d'eau : assainissement, gestion de...	85

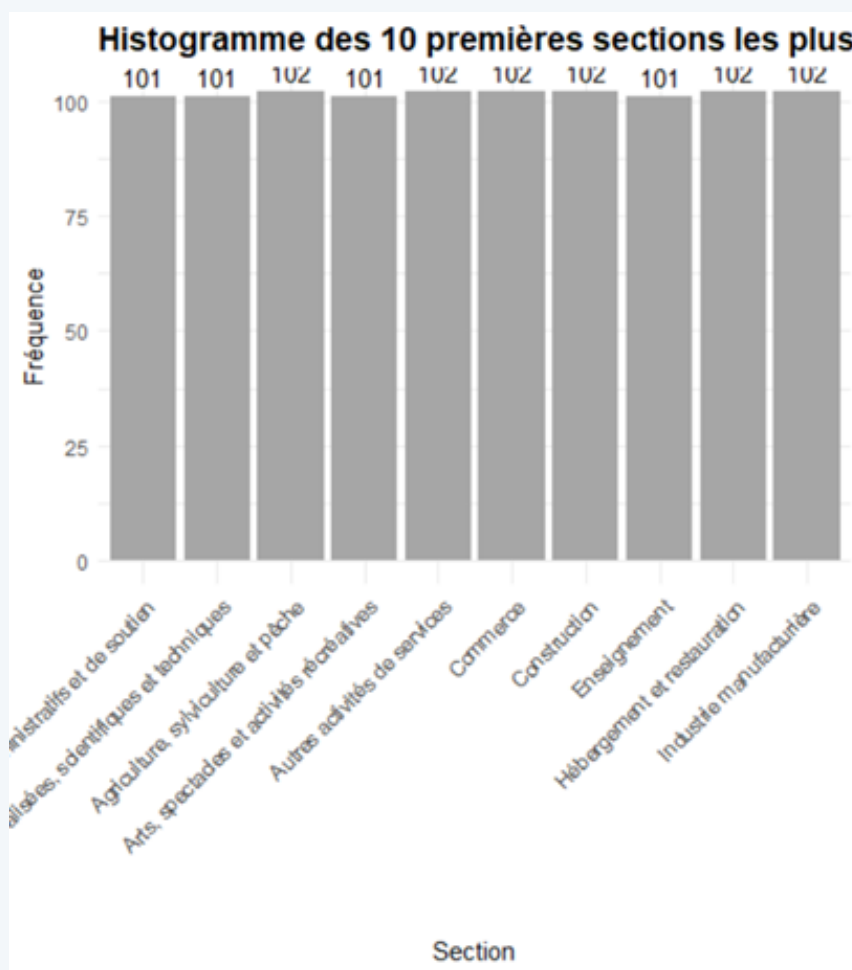
L'analyse de l'histogramme des données confirme une faible dispersion entre les départements. En effet, nous observons une distribution relativement uniforme des fréquences, ce qui indique une similarité marquée dans les valeurs entre les différents départements. Cette constatation suggère que les prêts accordés pendant la période étudiée sont répartis de manière équilibrée et que les variations entre les départements sont minimales.

Pour améliorer la qualité de notre histogramme, nous avons opté pour l'utilisation de la librairie `ggplot2`.

Cette puissante librairie offre des fonctionnalités avancées pour la création de graphiques, nous permettant ainsi de personnaliser facilement l'apparence de notre histogramme. De plus, nous avons ajouté des étiquettes à l'aide de la fonction `geom_text()` de `ggplot2`, ce qui nous permet d'obtenir un aperçu plus clair de la dispersion des données. Grâce à ces ajustements, nous sommes en mesure de visualiser plus efficacement les fréquences et d'identifier toute tendance ou variation significative dans notre analyse.

VARIABLES QUALITATIVES :

Histogramme :



Lors de notre analyse, nous avons identifié les sections d'activités qui ont été les plus représentées en termes d'aides accordées. Parmi ces sections, l'industrie manufacturière, l'hébergement et la restauration, la construction, le commerce, les autres activités de services, ainsi que l'agriculture, la sylviculture et la pêche se sont démarquées. Ces secteurs ont bénéficié d'un nombre important d'aides, ce qui souligne leur importance dans les mesures de soutien mises en place pendant la période étudiée. Ces résultats permettent de mettre en lumière les domaines économiques qui ont été les plus touchés par les conséquences de la pandémie de COVID-19 et qui ont donc bénéficié d'un soutien financier significatif. Cette analyse révèle ainsi les secteurs-clés qui ont été au cœur des efforts de relance et d'accompagnement dans le contexte de la crise sanitaire.

ANALYSES STATISTIQUES

BIVARIÉ :

L'analyse statistique bivariée, contrairement à l'analyse univariée, étudie la relation entre deux variables simultanément dans un ensemble de données. Elle permet d'explorer la nature de la relation entre ces variables, qu'elles soient de nature quantitative ou qualitative. L'analyse bivariée peut être utilisée pour déterminer s'il existe une corrélation, une association ou une dépendance entre les variables, ainsi que pour identifier les schémas ou les tendances spécifiques qui peuvent émerger de cette relation. Elle implique l'utilisation de diverses techniques statistiques telles que les tests d'indépendance, les régressions, les analyses de variance, etc. L'analyse bivariée permet d'approfondir la compréhension des relations entre les variables dans un contexte plus complexe, ce qui peut fournir des informations plus riches et une meilleure interprétation des données.

Nous allons lier les deux variables qualitatives avec les trois variables quantitatives. Ceci nous donne plusieurs couples de variables :

- Nombre de Prêt / Section
- Montant Total / Section
- Nombre de Prêt / Région
- Montant Total / Région
- Nombre de Prêt / Département
- Montant Total / Département

Nous avons utilisé la même méthode pour chaque couple de variable.

Nous allons détailler pour le premier couple afin de ne pas avoir de répétition.

Nous allons commencer par les bivariées liées à la variable quantitatives Section :

Nombre de Prêt par Section :

Dans un premier temps, nous avons utilisé la fonction `aggregate()` avec l'option `sum` pour obtenir le nombre total de prêts pour chaque section. Cela nous a permis d'agréger les données en fonction de la section et de calculer la somme des prêts correspondante. Ainsi, nous avons obtenu une nouvelle table qui présente le nombre total de prêts pour chaque section.

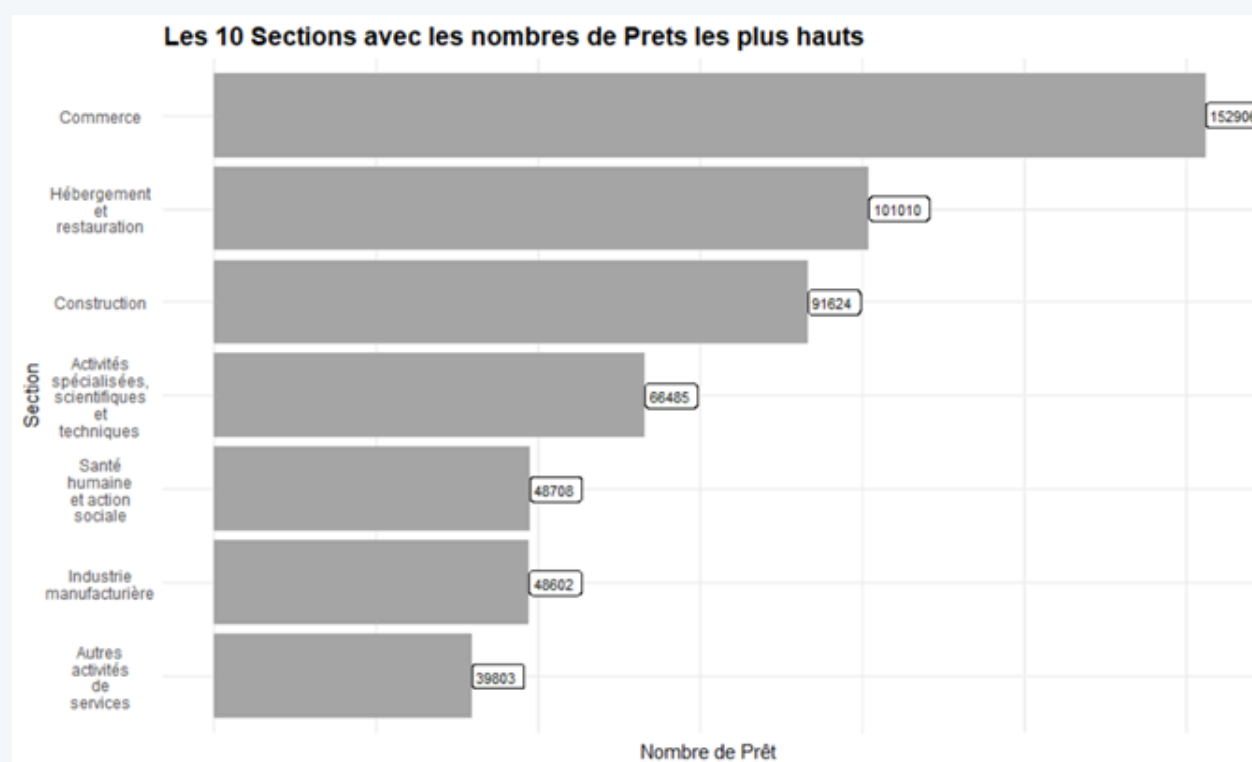
La fonction `aggregate()` est utilisée pour regrouper et résumer les données en fonction d'une ou plusieurs variables spécifiées. Elle prend en entrée un ensemble de données et permet de calculer des statistiques pour différentes sous-catégories définies par ces variables. Par exemple, en utilisant l'option `sum`, on peut regrouper les données selon une variable et obtenir la somme des valeurs correspondantes pour chaque groupe. Cela permet d'obtenir un résumé des données en agrégeant les informations de manière significative.

Nombre de Prêt par Section :

Nous avons créé la table Bivarie_TR qui regroupe la somme par Section du montant Total.

	Base_de_Donnée\$Nom_Section	Base_de_Donnée\$Nombre_de_Pret
9	Commerce	152906
12	Hébergement et restauration	101010
10	Construction	91624
4	Activités spécialisées, scientifiques et techniques	66485
19	Santé humaine et action sociale	48708
13	Industrie manufacturière	48602
8	Autres activités de services	39803
1	Activités de services administratifs et de soutien	25786
20	Transports et entreposage	23879
6	Agriculture, sylviculture et pêche	23385
15	Information et communication	18018
3	Activités immobilières	17144
7	Arts, spectacles et activités récréatives	14693
11	Enseignement	13335
2	Activités financières et d'assurance	11711
17	Production et distribution d'eau ; assainissement, gestion de...	1372
16	Non connu	686

Pour avoir un histogramme pertinent, nous avons ordonné les données dans l'ordre décroissant



La section qui a eu le plus de prêt garanti par l'Etat est la section Commerce avec 152 906.

Montant Total par Section :

Nous avons créé la table Bivarie_TS qui regroupe la somme par Section du montant Total.

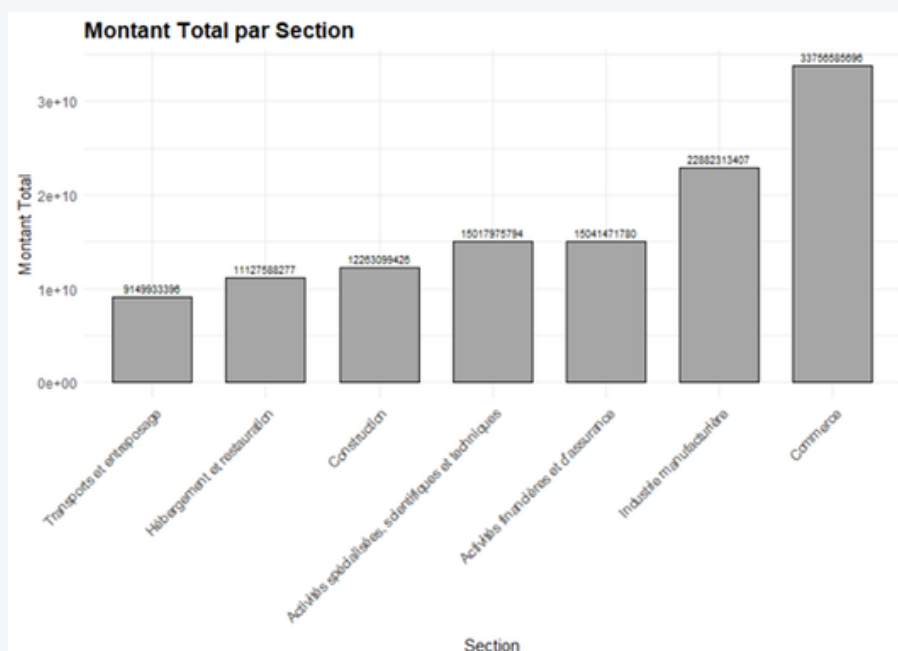
Montant Total par Section :

Nous avons créé la table Bivarie_TR qui regroupe la somme par Section du montant Total.

	Base_de_Donnée\$Nom_Section	Base_de_Donnée\$Montant_Total
9	Commerce	33756585696
13	Industrie manufacturière	22882313407
2	Activités financières et d'assurance	15041471780
4	Activités spécialisées, scientifiques et techniques	15017975794
10	Construction	12263099426
12	Hébergement et restauration	11127588277
20	Transports et entreposage	9149933396

Le commerce se démarque à nouveau en enregistrant le montant total le plus élevé, atteignant plus de 33 milliards.

Histogramme :



Ce graphique souligne une disparité considérable entre le secteur du commerce et les autres secteurs, en particulier l'écart important avec le deuxième secteur le plus important, à savoir l'industrie manufacturière. Il est remarquable que le montant total du secteur du commerce soit trois fois supérieur à celui du septième secteur, à savoir les transports et l'entreposage.

Résumé :

Avec l'analyse bivariée des deux couples en fonction de la région et des critères, à savoir la section et le montant total des prêts garantis par l'État, il est clair que la section du Commerce se démarque nettement des autres sections. En effet, elle enregistre à la fois le plus grand nombre de prêts garantis par l'État et le montant total le plus élevé, surpassant ainsi les autres sections de manière significative.

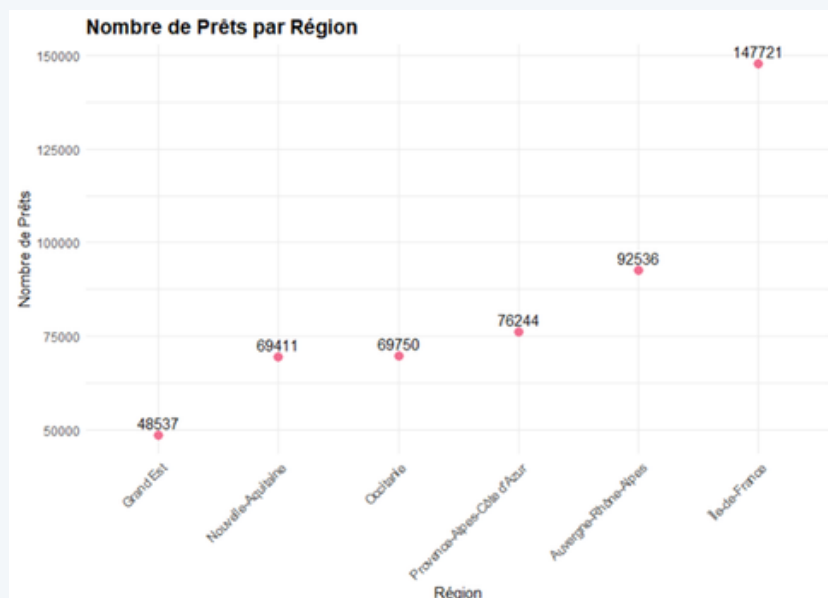
Nombre de Prêt par Région :

Nous avons créé la table Bivarie_PR qui regroupe la somme par Section du montant Total.

	Nom_Region	Nombre_de_Pret
10	Île-de-France	147721
1	Auvergne-Rhône-Alpes	92536
18	Provence-Alpes-Côte d'Azur	76244
16	Occitanie	69750
15	Nouvelle-Aquitaine	69411
6	Grand Est	48537

L'Île-de-France est la région qui a bénéficié du plus grand nombre de Prêts Garantis par l'État, avec un total de 147,721 prêts accordés. Cette région, qui comprend notamment Paris, a joué un rôle majeur dans le soutien financier aux entreprises pendant cette période. Ces prêts ont contribué à stimuler l'économie régionale et à soutenir les entreprises face aux difficultés rencontrées en raison de la pandémie ou d'autres circonstances exceptionnelles.

Afin de pouvoir observer la dispersion en fonction des régions nous avons fait un graphique.



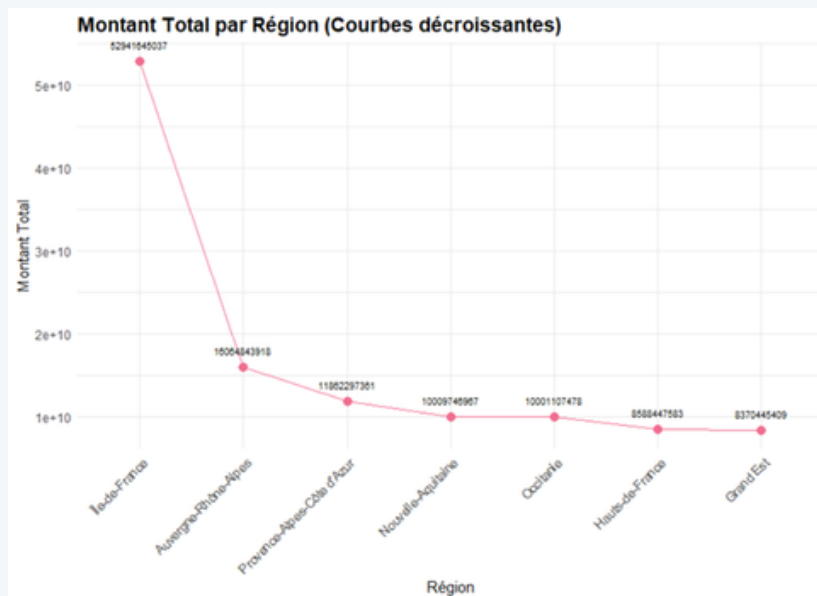
Le graphique met en évidence un écart significatif entre la région Île-de-France et les autres régions, notamment en termes de nombre de prêts accordés. L'Île-de-France se démarque nettement en bénéficiant du plus grand nombre de Prêts Garantis par l'État, avec un écart de plus de 50 000 prêts par rapport à la deuxième région, l'Auvergne-Rhône-Alpes. Cette disparité souligne l'importance économique de la région parisienne et son rôle clé dans le soutien financier aux entreprises. Les prêts accordés ont été essentiels pour aider les entreprises de l'Île-de-France à faire face aux défis économiques, en particulier ceux engendrés par la pandémie.

Montant Total par Région :

Nous avons créé la table Bivarie_TR qui regroupe la somme par Section du montant Total.

	Nom_Région	Montant_Total
10	Île-de-France	52941645037
1	Auvergne-Rhône-Alpes	16064843918
18	Provence-Alpes-Côte d'Azur	11862297361
15	Nouvelle-Aquitaine	10009746967
16	Occitanie	10001107478
9	Hauts-de-France	8588447583
6	Grand Est	8370445409

Il existe un écart important, une fois de plus, entre l'Île-de-France et les autres régions. Même si nous excluons l'Île-de-France, nous pouvons constater qu'il y a toujours un écart significatif entre les régions, notamment entre l'Auvergne-Rhône-Alpes qui a un montant total deux fois plus élevé que la région Grand Est.



Résumé :

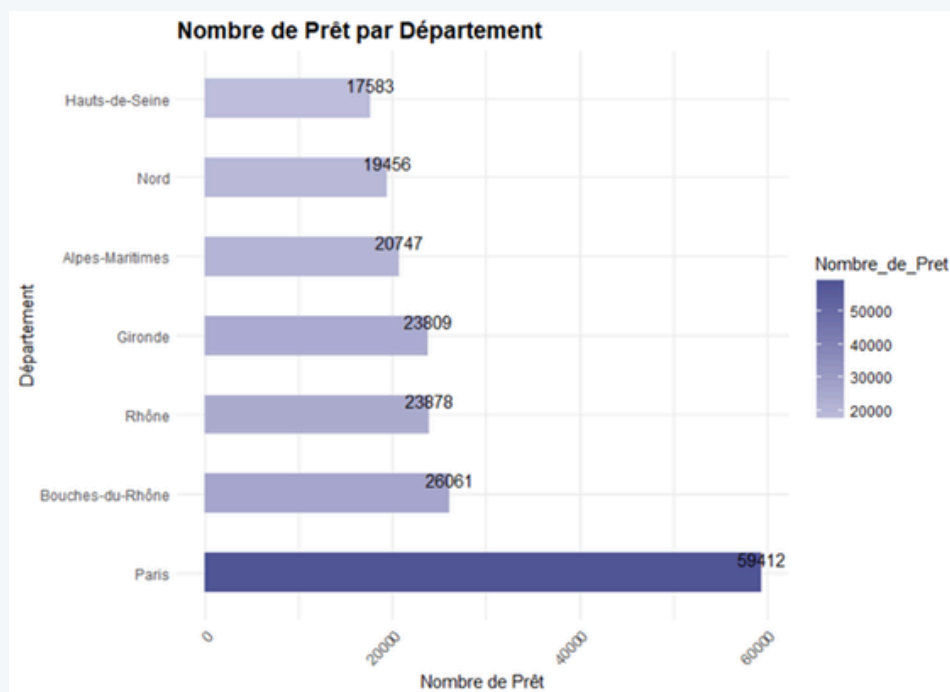
Avec l'analyse bivariee des variables liées à la région et aux critères du nombre de prêts garantis par l'État et du montant total, il est évident que la région qui se distingue en termes du plus grand nombre de prêts et du montant total est l'Île-de-France.

Nombre de Prêt par département :

Nous avons créé la table Bivarie_PD qui regroupe la somme par Section du montant Total.

	Nom_Département	Nombre_de_Pret
77	Paris	59412
13	Bouches-du-Rhône	26061
82	Rhône	23878
34	Gironde	23809
5	Alpes-Maritimes	20747
74	Nord	19456
47	Hauts-de-Seine	17583

Paris est le département qui a reçu le plus de prêts garantis par l'État, notamment en raison de son statut de capitale. Nous pouvons constater un écart significatif entre Paris et les autres départements, en particulier avec les Bouches-du-Rhône, qui ont reçu seulement la moitié du nombre de prêts par l'État comparé à Paris.



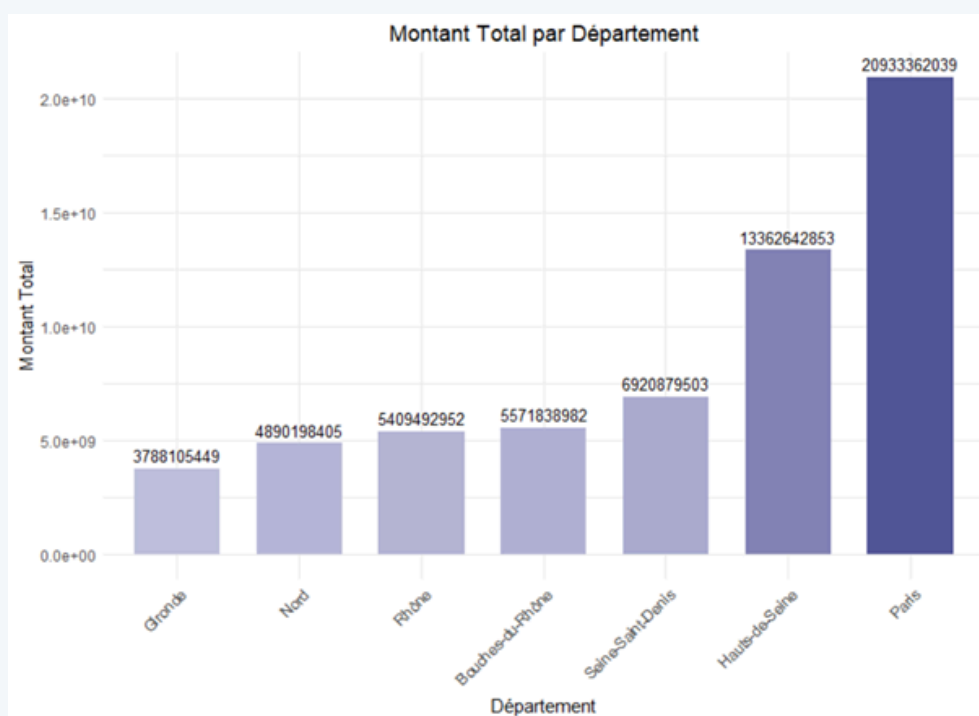
L'observation de ce graphique met clairement en évidence l'écart entre les départements et Paris. En revanche, parmi les autres départements, nous constatons que l'écart n'est pas aussi prononcé.

Montant Total par Département :

Nous avons créé la table Bivarie_TD qui regroupe la somme par Section du montant Total.

	Nom_Département	Montant_Total
77	Paris	20933362039
47	Hauts-de-Seine	13362642853
88	Seine-Saint-Denis	6920879503
13	Bouches-du-Rhône	5571838982
82	Rhône	5409492952
74	Nord	4890198405
34	Gironde	3788105449

En examinant ce couple de variables, nous pouvons constater que Paris se distingue une fois de plus des autres départements, bien que dans une moindre mesure que dans le couple de variables Nombre de Prêt/Nom de Région. Cela suggère que Paris a reçu un grand nombre de prêts avec un montant total relativement bas. De plus, nous pouvons également remarquer que les Hauts-de-Seine se démarquent également du reste du groupe inférieur.



Résumé de l'analyse Bivarié :

Nous avons analysé un total de six couples de variables, ce qui nous a apporté de nombreuses précisions sur les inégalités de répartition entre les sections, les régions et les départements. En ce qui concerne les sections, nous pouvons clairement identifier la section Commerce comme étant celle qui se démarque des autres avec le plus grand montant total.

La dispersion entre les différentes sections met en évidence les fortes inégalités en matière de Prêts Garantis par l'État.

Concernant les régions, une fois de plus, nous observons une grande inégalité entre elles, avec la région Île-de-France qui domine largement les autres. Cette région se distingue par un montant total bien supérieur aux autres régions.

En ce qui concerne les départements, sans surprise, c'est la capitale, Paris, qui a bénéficié du plus grand nombre d'aides et de prêts de la part de l'État, la devançant ainsi tous les autres départements.

Corrélation :

Nous allons nous concentrer sur un autre couple de variables : nous allons calculer plusieurs indicateurs afin de déterminer s'il existe une corrélation entre les variables `Montant_Total` et `Nombre_de_Pret`.

En statistique, la corrélation désigne la mesure de la relation linéaire entre deux variables. Elle permet de déterminer dans quelle mesure les variations d'une variable sont associées aux variations d'une autre variable.

Grâce à la fonction `cor()` qui nous a permis de calculer la corrélation entre les deux valeurs

```
> cor(Base_de_Donnée$Nombre_de_Pret,Base_de_Donnée$Montant_Total)
[1] 0.6601348
```

Une corrélation de 0,66 indique une corrélation positive, relativement forte entre les deux variables étudiées. Cela signifie qu'il existe une tendance claire où lorsque la valeur d'une variable augmente, la valeur de l'autre variable a tendance à augmenter également.

Cela nous montre donc que si le nombre de Prêt Augment alors le montant total augmente aussi.

RÉGRESSION LINÉAIRE :

La régression linéaire est une technique statistique utilisée pour modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes. Elle cherche à établir une relation linéaire entre ces variables afin de prédire la valeur de la variable dépendante en fonction des valeurs des variables indépendantes.

Le but de faire une régression dans notre contexte, c'est de montrer au mieux l'inégalité des Prêts.

Nous avons donc fait une liaison encore une fois des deux variables quantitatives, montant total et Nombre de Prêt.

Nous avons assigné :

X : Au Nombre de prêts

Y : Au Montant Total

Nous avons utilisé plusieurs modèles avec des degrés différents afin de pouvoir expliquer au mieux la dispersion.

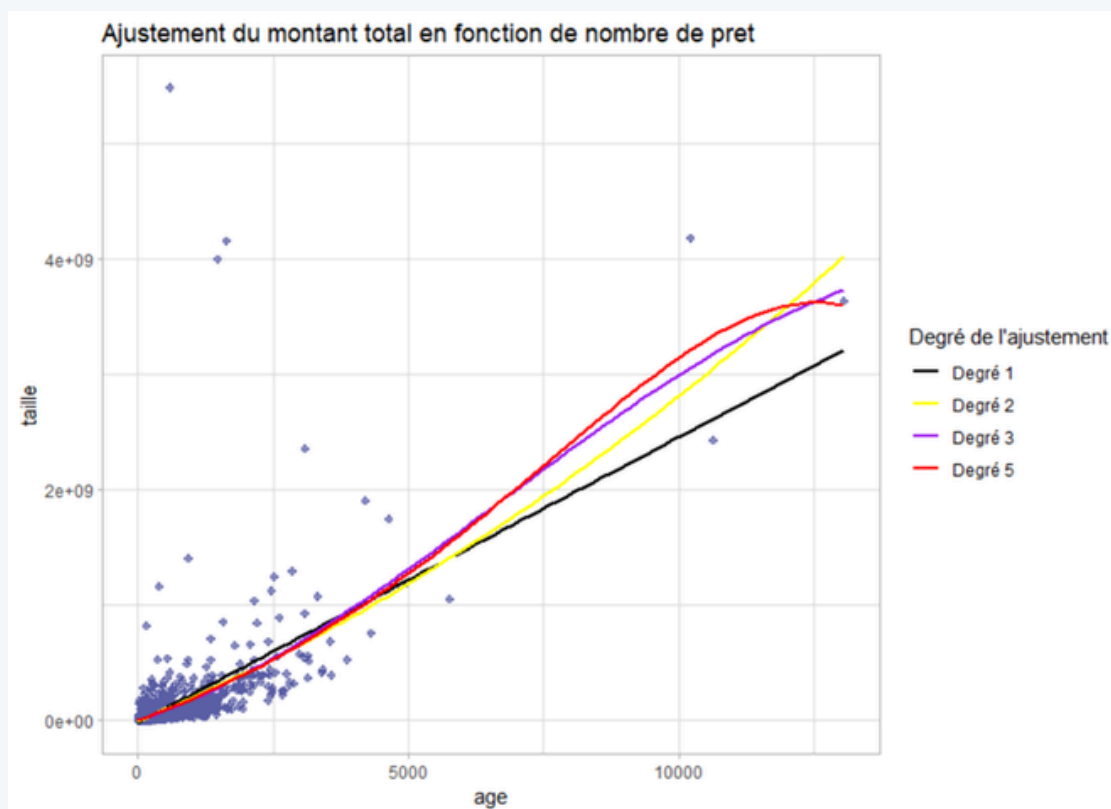
Nous les avons représentées sur un graphique.

Avant ceci, le calcul du R^2 nous permet d'avoir un aperçu de l'ajustement le plus judicieux.

Plus le R^2 est élevé plus l'ajustement peut expliquer au mieux la dispersion des données.

Voici le tableau qui nous montre le R^2 de chaque model :

Model Degré 1	Model Degré 2	Model Degré 3	Model Degré 5	Model Degré Max
0.44	0.45	0.45	0.451	0.451



Nous pouvons observer des indications à partir du nuage de points qui illustre la relation entre le montant total et le nombre de prêts. Il est clair que le nuage de points ne présente pas une forme homogène. Certains points se trouvent à des valeurs élevées sur l'axe du montant total, tandis que le nombre de prêts correspondant est faible. Cela suggère qu'il existe des cas où le montant total est élevé sans nécessairement avoir un nombre de prêts élevé. En d'autres termes, il y a une certaine variation dans la relation entre le montant total et le nombre de prêts, et certains individus peuvent avoir un montant total important malgré un nombre de prêts relativement faible.

CONCLUSION

Comme nous l'avons constaté dans notre étude, les prêts garantis par l'État sont des outils indispensables pour les organisations telles que les sections, les régions et les départements afin de se relancer après la pandémie de Covid-19. Les prêts accordés par l'État ont pour objectif d'aider ces organisations spécifiques.

Nous avons pu observer une répartition inégale de ces prêts selon les sections, les régions et les départements. Grâce à l'étude de nos variables, tant en analyse bivariable qu'en analyse univariée, nous avons pu dégager certaines tendances. Nous avons constaté que certaines sections, régions et départements sont particulièrement favorisés et reçoivent des prêts de l'État d'un montant élevé.

Dans la grande majorité des cas, nous avons observé une cohérence entre l'analyse statistique univariée et l'analyse bivariable. Par exemple, l'analyse univariée a révélé que Paris est l'un des départements les plus représentés, et cela a été confirmé par le fait que Paris a reçu le plus de prêts garantis par l'État, avec un montant total bien plus élevé que les autres départements.

En ce qui concerne l'analyse univariée des sections, nous avons constaté que le commerce est l'une des sections les plus représentées, et c'est également la section qui a reçu le plus de prêts de l'État, avec un montant total élevé.

Pour l'analyse univariée des départements, nous avons observé que ce n'est pas nécessairement le département le plus représenté qui a reçu le plus de prêts de l'État et le montant total le plus élevé. Par exemple, bien que l'Occitanie soit le département le plus représenté, c'est l'Île-de-France qui a reçu le plus de prêts avec le montant le plus élevé.

La régression nous a permis de confirmer nos observations à travers l'analyse du nuage de points et des modèles que nous avons créés. Le nuage de points nous montre que certains montants totaux sont élevés alors que le nombre de prêts associés à ces montants est faible. Cependant, nous ne disposons pas d'indications sur les facteurs politiques, sociaux et démographiques pouvant influencer le choix de l'État de favoriser certaines sections, départements ou régions.

Notre étude met en évidence des inégalités sans fournir d'explications sur les facteurs qui pourraient les influencer.



MERCI