

TRAVAIL PRATIQUE

COURS D'UNIVERSITÉ
ABC-1234

ÉQUIPE 55

Rapport Nom de votre TP

Par

Étudiant 1
Étudiant 2
Étudiant 3
Étudiant 4

Numéro d'identification

XYZ XYZ XYZ
YZX YZX YZX
ZYX ZYX ZYX
XYD XYD XYD

*Travail présenté à
Madame*

PROF D'UNIVERSITÉ

32 AVRIL 2025



UNIVERSITÉ
LAVAL

Faculté des sciences et de génie
École d'actuariat

Table des Matières

Titre 1	2
Sous-titre 1	2
Sélection des variables	2
Création de la nouvelle variable réponse	2
Titre 2	2

Titre 1

Sous-titre 1

Sélection des variables

La première étape du travail a consisté à réduire la dimension du jeu de données. En effet, celui-ci est constitué de 41 variables, dont une bonne partie n'étant pas utiles dans le contexte de l'analyse des montants de réclamation.

Sans effectuer aucune analyse statistique, nous avons jugé adéquat de retirer plusieurs variables du modèle, notamment, toutes les variables contenant beaucoup de valeurs manquantes, comme `baseFloodElevation`, `basementEnclosureCrawlspace`, `elevationCertificateIndicator`, `elevationDifference`, `rateMethod` et `lowestAdjacentGrade`. Ces variables sont aussi toutes issues de l'évaluation de quelques uns des bâtiments assurés, alors que plusieurs autres variables telles que `numberOfFloorsInTheInsuredBuilding`, `originalConstructionDate` ou encore `lowestFloorElevation` auront un impact probablement plus marqué sur le modèle sans devoir nécessiter un travail ardu et approximatif d'estimation d'une grande quantité de données manquantes.

Nous avons aussi pris la décision d'enlever les variables temporelles à l'exception de la date de construction du bâtiment (`originalConstructionDate`) et la date du sinistre (`dateOfLoss`), puisqu'elles sont les seules variables temporelles pertinentes à notre analyse selon nous.

Création de la nouvelle variable réponse

Dans le jeu de données se retrouvent trois colonnes contenant des informations sur les montants de prestations payés en lien avec le bâtiment (`amountPaidOnBuildingClaim`), les biens (`amountPaidOnContentsClaim`) et l'augmentation des coûts en lien avec la conformité (`amountPaidOnIncreasedCostOfComplianceClaim`).

Titre 2

```
data.raw <- read.csv("Flood_California.csv")

## Retirer les variables inutiles
data.rm <- data.raw[, c(1, 3, 4, 5, 6, 13, 14, 15, 16, 21, 25, 28, 33, 39, 41)]
data <- data.raw[, -c(1, 3, 4, 5, 6, 13, 14, 15, 16, 21, 25, 28, 33, 39, 41)]

# Combiner les variables réponses (totalAmount)
data$amountPaidOnBuildingClaim[is.na(data$amountPaidOnBuildingClaim)] <- 0
data$amountPaidOnBuildingClaim <-
  abs(data$amountPaidOnBuildingClaim)
data$amountPaidOnContentsClaim[is.na(data$amountPaidOnContentsClaim)] <- 0
data$amountPaidOnContentsClaim <-
  abs(data$amountPaidOnContentsClaim)
data$amountPaidOnIncreasedCostOfComplianceClaim[is.na(data$amountPaidOnIncreasedCostOfComplianceClaim)] <- 0
data$amountPaidOnIncreasedCostOfComplianceClaim <-
  abs(data$amountPaidOnIncreasedCostOfComplianceClaim)
data$totalAmount <- apply(data[, 17:19], 1, sum)
data <- data[, -c(17, 18, 19)]

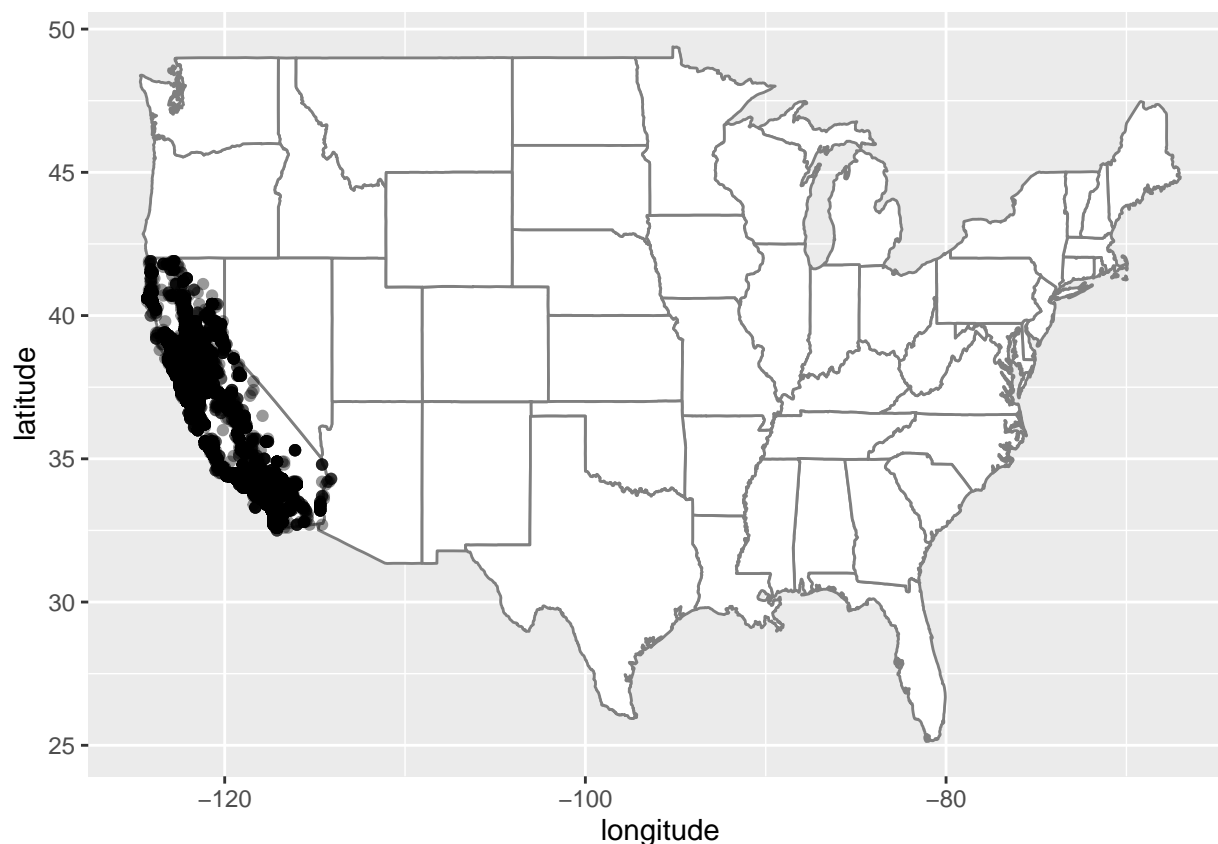
# Retirer les lignes n'étant pas localisées en Californie
data <- data[!is.na(data$longitude),]
data <- data[data$longitude <= -110,]
```

```
xdf <- which(is.na(data$countyCode))

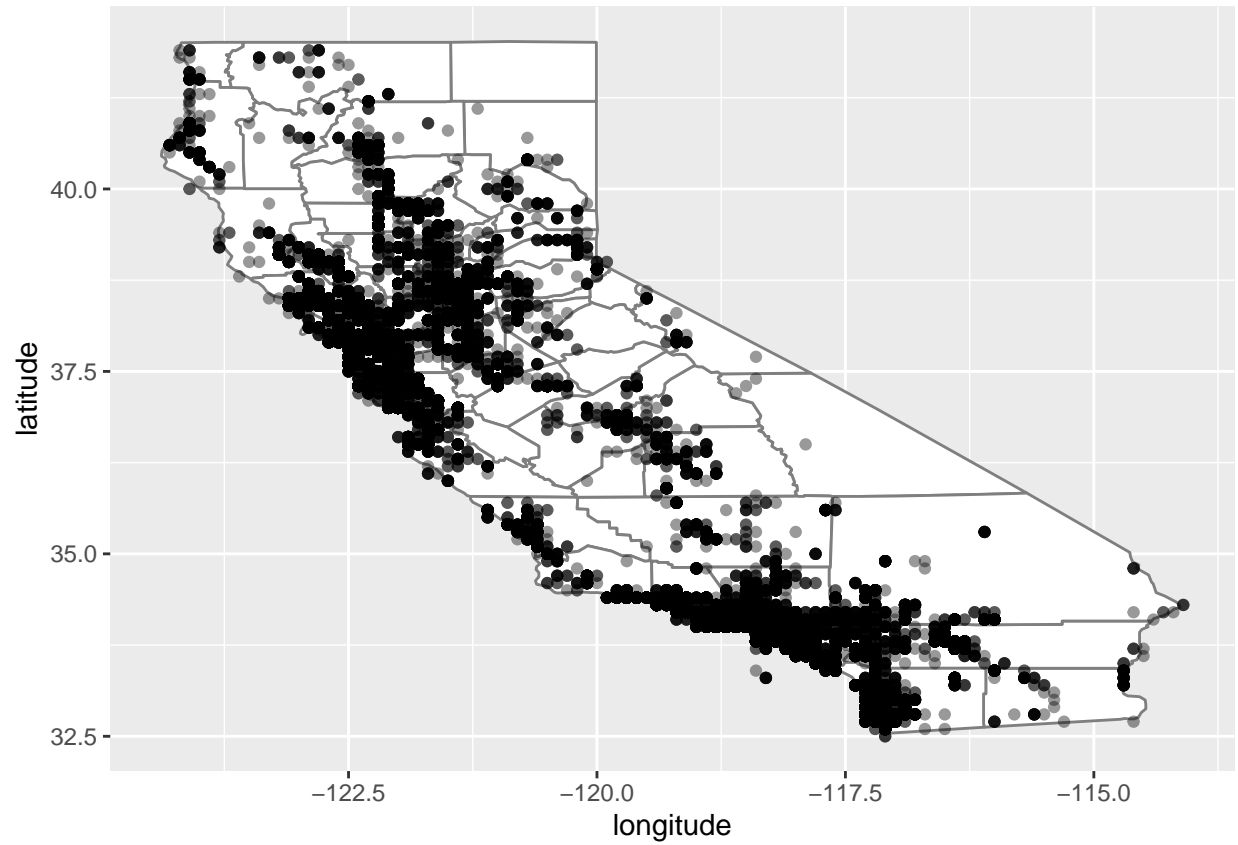
# Imputation par régression linéaire des codes de régions (countyCode)
mod.county <- lm(countyCode ~ latitude + longitude, data = data)
pred.county <- predict(mod.county, newdata = data[is.na(data$countyCode),], type = "response")
data$countyCode[is.na(data$countyCode)] <- pred.county
data <- data[data$countyCode != 32031,]

# Imputation par régression linéaire des codes de régions (countyCode)
# data$communityRatingSystemDiscount <- as.factor(data$communityRatingSystemDiscount)
# levels(data$communityRatingSystemDiscount) <- c("A", "B", "C", "D", "E", "F", "G", "H", "I", "H")
# mod.communityRating <- lm(communityRatingSystemDiscount ~ ., data = data)
# pred.CR <- predict(mod.communityRating, newdata = data[is.na(data$communityRatingSystemDiscount),], type = "response")
# data$communityRatingSystemDiscount[is.na(data$communityRatingSystemDiscount)] <- pred.CR

mapUSA <- borders(database = "state",
                  colour="gray50", fill="white")
ggplot(data = data, aes(x = longitude, y = latitude)) +
  mapUSA + geom_point(alpha = .4)
```



```
mapCalifornia <- borders(database = "county", region = "california",
                        colour="gray50", fill="white")
ggplot(data = data, aes(x = longitude, y = latitude)) +
  mapCalifornia + geom_point(alpha = .4)
```



```
md.pattern(data, rotate.names = T)
```

	agricultureStructureIndicator	condominiumIndicator	policyCount	countyCode	dateOfLoss	elevatedBuildingIndicator	houseWorship	latitude	longitude	locationOfContents	lowestFloorElevation	nonProfitIndicator	originalConstructionDate	postFIRMConstructionIndicator	smallBusinessIndicatorBuilding	state	totalBuildingInsuranceCoverage	totalContentsInsuranceCoverage	yearOfLoss	primaryResidence	totalAmount	occupancyType	numberOfFloorsInTheInsuredBuilding	communityRatingSystemDiscount
36387	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14275	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1930420645		
##	agricultureStructureIndicator condominiumIndicator policyCount countyCode																							
## 36387										1							1				1		1	
## 14275										1							1				1		1	
## 16										1							1				1		1	
## 12										1							1				1		1	
## 10										1							1				1		1	
## 7										1							1				1		1	
## 2										1							1				1		1	
##										0							0				0		0	
##	dateOfLoss elevatedBuildingIndicator houseWorship latitude longitude																							
## 36387					1		1		1								1				1		1	
## 14275					1		1		1								1				1		1	
## 16					1		1		1								1				1		1	
## 12					1		1		1								1				1		1	
## 10					1		1		1								1				1		1	
## 7					1		1		1								1				1		1	
## 2					1		1		1								1				1		1	
##					0		0		0								0				0		0	
##	locationOfContents lowestFloorElevation nonProfitIndicator																							
## 36387						1					1										1			
## 14275						1					1										1			
## 16						1					1										1			
## 12						1					1										1			
## 10						1					1										1			
## 7						1					1										1			
## 2						1					1										1			

##	0	0	0
##	originalConstructionDate	postFIRMConstructionIndicator	
## 36387	1		1
## 14275	1		1
## 16	1		1
## 12	1		1
## 10	1		1
## 7	1		1
## 2	1		1
##	0		0
##	smallBusinessIndicatorBuilding	state	totalBuildingInsuranceCoverage
## 36387	1	1	1
## 14275	1	1	1
## 16	1	1	1
## 12	1	1	1
## 10	1	1	1
## 7	1	1	1
## 2	1	1	1
##	0	0	0
##	totalContentsInsuranceCoverage	yearOfLoss	primaryResidence
## 36387	1	1	1
## 14275	1	1	1
## 16	1	1	1
## 12	1	1	1
## 10	1	1	1
## 7	1	1	1
## 2	1	1	1
##	0	0	0
##	occupancyType	numberOfFloorsInTheInsuredBuilding	
## 36387	1		1
## 14275	1		1
## 16	1		0
## 12	1		0
## 10	0		1
## 7	0		1
## 2	0		0
##	19		30
##	communityRatingSystemDiscount		
## 36387	1	0	
## 14275	0	1	
## 16	1	1	
## 12	0	2	
## 10	1	1	
## 7	0	2	
## 2	0	3	
##	14296	14345	