Datos

Test

Lourdes Sofía Elizondo Guajardo — Daniela Diaz Delgado — Gabriel González Bataller

Bernardo Ortega Chávez

2021-06-21

Contents

| 1 | Dat | ta Analysis | 2 |
|---|-----|--|----|
| | 1.1 | Data Reported in Sistema de Información Infonavit: $2013/01-2021/04$ | 2 |
| | | 1.1.1 Number of Credits Granted | 3 |
| | | 1.1.2 Credit Amount (MXN) | 13 |
| | | 1.1.3 Subsidies | 22 |
| | | 1.1.3.1 Number | 25 |
| | | 1.1.3.2 Amount (\$MXN) | 33 |
| | 1.2 | Data Reported in Sistema de Información Infonavit: Complete Portfolio Analysis | 39 |
| 2 | San | mple data | 40 |
| | 2.1 | Data Analysis | 40 |
| | 2.2 | Adjust Sample to Match Population's Characteristics | 42 |
| | 2.3 | Asign Variable Status | 42 |
| 3 | Cre | edit Risk Model | 45 |
| | 3.1 | Logistic Regression | 43 |
| | 3.2 | Evaluation | 43 |
| | 3.3 | Regional Analysis & Clusters | 43 |
| 4 | Bui | ilding a Cross-Subsidy Scheme | 43 |
| | 4.1 | Asigning an Interest Rate | 43 |
| 5 | Cor | mparative Statics & Tests | 4: |

1 Data Analysis

1.1 Data Reported in Sistema de Información Infonavit: 2013/01-2021/04

Information regarding all mortgages originated by Infonavit from January 2013 to April 2021.

 $\label{lem:https://portalmx.infonavit.org.mx/wps/portal/infonavit.web/el-instituto/el-infonavit/SII/!ut/p/z1/pZBBDoIwFETP4gn-FJHWJRosFQKiqUA3pitCoujCeH7RJcSKcXY_eTN_MmSoItPZR9vYe3vt7Lm_axOcAgnEa59lQh05iijwc873yAWncgCkyQIFdonEFuzAQOYfv_am-T0ByBVYJkUsEG4iFS3TuSc1m-bHB4U_9x8Bxh1fkhm8GC_oBF4TvQHXBt9a3C66V4VWNbMnVfmrRQ!!/dz/d5/L2dBISEvZ0FBIS9nQSEh/$

```
# Número de Créditos Formalizados
NCTotal <- as.data.frame(read_excel("DATA/NumeroDeCreditosFormalizados.xlsx", sheet = 1))
MCTipoProducto<-as.data.frame(read_excel("DATA/NumeroDeCreditosFormalizados.xlsx", sheet = 2))
NCLinea <- as.data.frame(read_excel("DATA/NumeroDeCreditosFormalizados.xlsx", sheet = 3))
NCClasificacionVivienda<-as.data.frame(read_excel("DATA/NumeroDeCreditosFormalizados.xlsx", sheet = 4))</pre>
NCIngreso<-as.data.frame(read_excel("DATA/NumeroDeCreditosFormalizados.xlsx", sheet = 5))</pre>
NCEdad <- as.data.frame(read_excel("DATA/NumeroDeCreditosFormalizados.xlsx", sheet = 6))
NCEstado<-as.data.frame(read_excel("DATA/NumeroDeCreditosFormalizados.xlsx", sheet = 7))
## Totales
NC_TP<-as.data.frame(read_excel("DATA/Totales/NC.xlsx",sheet=1))</pre>
NC_L<-as.data.frame(read_excel("DATA/Totales/NC.xlsx",sheet=2))</pre>
NC_CV<-as.data.frame(read_excel("DATA/Totales/NC.xlsx",sheet=3))</pre>
NC I <- as.data.frame(read excel("DATA/Totales/NC.xlsx",sheet=4))
NC_Edad<-as.data.frame(read_excel("DATA/Totales/NC.xlsx",sheet=5))</pre>
NC_Estado<-as.data.frame(read_excel("DATA/Totales/NC.xlsx",sheet=6))</pre>
# Importe de Cheque
ITotal<-as.data.frame(read_excel("DATA/ImporteDeCheque.xlsx", sheet = 1))</pre>
ITipoProducto<-as.data.frame(read_excel("DATA/ImporteDeCheque.xlsx", sheet = 2))</pre>
ILinea<-as.data.frame(read_excel("DATA/ImporteDeCheque.xlsx", sheet = 3))</pre>
IClasificacionVivienda<-as.data.frame(read_excel("DATA/ImporteDeCheque.xlsx", sheet = 4))</pre>
IIngreso<-as.data.frame(read_excel("DATA/ImporteDeCheque.xlsx", sheet = 5))</pre>
IEdad<-as.data.frame(read_excel("DATA/ImporteDeCheque.xlsx", sheet = 6))</pre>
IEstado<-as.data.frame(read_excel("DATA/ImporteDeCheque.xlsx", sheet = 7))</pre>
## Totales
I TP<-as.data.frame(read excel("DATA/Totales/I.xlsx",sheet=1))
I_L<-as.data.frame(read_excel("DATA/Totales/I.xlsx",sheet=2))</pre>
I_CV<-as.data.frame(read_excel("DATA/Totales/I.xlsx",sheet=3))</pre>
I_I<-as.data.frame(read_excel("DATA/Totales/I.xlsx",sheet=4))</pre>
I_Edad<-as.data.frame(read_excel("DATA/Totales/I.xlsx",sheet=5))</pre>
I Estado<-as.data.frame(read excel("DATA/Totales/I.xlsx",sheet=6))
# Monto de Crédito Infonavit
MCTotal <- as.data.frame(read_excel("DATA/MontoDeCreditoInfonavit.xlsx", sheet = 1))
MCTipoProducto<-as.data.frame(read_excel("DATA/MontoDeCreditoInfonavit.xlsx", sheet = 2))</pre>
MCLinea <- as.data.frame(read_excel("DATA/MontoDeCreditoInfonavit.xlsx", sheet = 3))
MCClasificacionVivienda<-as.data.frame(read_excel("DATA/MontoDeCreditoInfonavit.xlsx", sheet = 4))</pre>
MCIngreso<-as.data.frame(read_excel("DATA/MontoDeCreditoInfonavit.xlsx", sheet = 5))</pre>
MCEdad<-as.data.frame(read_excel("DATA/MontoDeCreditoInfonavit.xlsx", sheet = 6))</pre>
MCEstado<-as.data.frame(read_excel("DATA/MontoDeCreditoInfonavit.xlsx", sheet = 7))</pre>
## Totales
```

```
MC_TP<-as.data.frame(read_excel("DATA/Totales/MC.xlsx",sheet=1))</pre>
MC_L<-as.data.frame(read_excel("DATA/Totales/MC.xlsx",sheet=2))</pre>
MC_CV<-as.data.frame(read_excel("DATA/Totales/MC.xlsx",sheet=3))</pre>
MC_I<-as.data.frame(read_excel("DATA/Totales/MC.xlsx",sheet=4))</pre>
MC_Edad<-as.data.frame(read_excel("DATA/Totales/MC.xlsx",sheet=5))</pre>
MC_Estado<-as.data.frame(read_excel("DATA/Totales/MC.xlsx",sheet=6))</pre>
# Número de Subsidios
NSTotal <- as.data.frame(read_excel("DATA/NumeroDeSubsidios.xlsx", sheet = 1))
NSTipoProducto<-as.data.frame(read_excel("DATA/NumeroDeSubsidios.xlsx", sheet = 2))
NSLinea<-as.data.frame(read_excel("DATA/NumeroDeSubsidios.xlsx", sheet = 3))
NSClasificacionVivienda<-as.data.frame(read_excel("DATA/NumeroDeSubsidios.xlsx", sheet = 4))
NSIngreso<-as.data.frame(read_excel("DATA/NumeroDeSubsidios.xlsx", sheet = 5))</pre>
NSEdad <- as.data.frame(read_excel("DATA/NumeroDeSubsidios.xlsx", sheet = 6))
NSEstado<-as.data.frame(read_excel("DATA/NumeroDeSubsidios.xlsx", sheet = 7))
## Totales
NS_TP<-as.data.frame(read_excel("DATA/Totales/NS.xlsx",sheet=1))</pre>
NS_L<-as.data.frame(read_excel("DATA/Totales/NS.xlsx",sheet=2))
NS_CV<-as.data.frame(read_excel("DATA/Totales/NS.xlsx",sheet=3))</pre>
NS_I<-as.data.frame(read_excel("DATA/Totales/NS.xlsx",sheet=4))
NS_Edad<-as.data.frame(read_excel("DATA/Totales/NS.xlsx",sheet=5))
NS_Estado<-as.data.frame(read_excel("DATA/Totales/NS.xlsx",sheet=6))
# Monto de Subsidios
MSTotal <- as.data.frame(read_excel("DATA/MontoDeSubsidios.xlsx", sheet = 1))
MSTipoProducto<-as.data.frame(read_excel("DATA/MontoDeSubsidios.xlsx", sheet = 2))
MSLinea<-as.data.frame(read_excel("DATA/MontoDeSubsidios.xlsx", sheet = 3))
MSClasificacionVivienda <- as.data.frame(read_excel("DATA/MontoDeSubsidios.xlsx", sheet = 4))
MSIngreso<-as.data.frame(read_excel("DATA/MontoDeSubsidios.xlsx", sheet = 5))
MSEdad<-as.data.frame(read_excel("DATA/MontoDeSubsidios.xlsx", sheet = 6))</pre>
MSEstado<-as.data.frame(read_excel("DATA/MontoDeSubsidios.xlsx", sheet = 7))
MS_TP<-as.data.frame(read_excel("DATA/Totales/MS.xlsx",sheet=1))
MS_L<-as.data.frame(read_excel("DATA/Totales/MS.xlsx",sheet=2))
MS_CV<-as.data.frame(read_excel("DATA/Totales/MS.xlsx",sheet=3))</pre>
MS_I<-as.data.frame(read_excel("DATA/Totales/MS.xlsx",sheet=4))
MS_Edad<-as.data.frame(read_excel("DATA/Totales/MS.xlsx",sheet=5))</pre>
MS_Estado<-as.data.frame(read_excel("DATA/Totales/MS.xlsx",sheet=6))
# Recaudación y Fiscalización de las Aportaciones Patronales
RFAP<-as.data.frame(read_excel("DATA/RecaudacionYFiscalizacionDeAportacionesPatronales.xlsx", sheet = 1
# Administración de la Subcuenta de Vivienda
ASV<-as.data.frame(read_excel("DATA/AdminSubcuentaDeVivienda.xlsx", sheet = 1))
```

1.1.1 Number of Credits Granted

```
# Por Estado
barplot_NC_Estado<-ggplot(data=NC_Estado)+
geom_bar(aes(x=reorder(Estado,-NC),y=NC,fill=Estado),stat = "identity")+</pre>
```

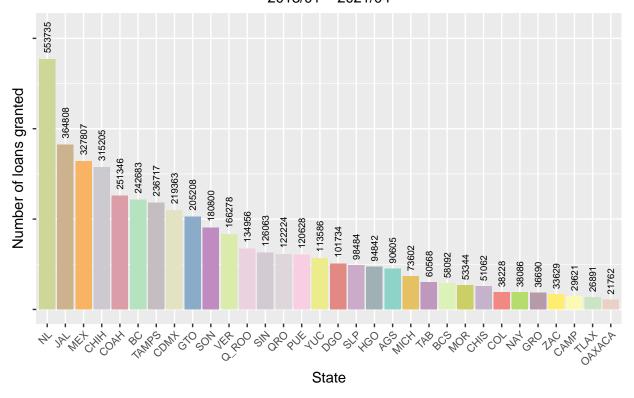
```
geom_text(aes(x=Estado,y=NC,label=NC),hjust=-0.2, size=2.5,angle=90)+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
  labs( x = "State",
        y = "Number of loans granted",
        title ="Loan distribution per State",
        subtitle = "2013/01 - 2021/04"
        )+
  ylim(0,625000) +
  theme(
    axis.text.x=element_text(size=8,angle=45,hjust=1),
    axis.text.y=element_text(size=0,angle=45),
    plot.title = element_text(hjust=0.5),
    plot.subtitle=element text(hjust=0.5),
    legend.position = "none"
# Por Estado como porcentaje
percent_Estado<-paste(round(100*(NC_Estado$NC/(sum(NC_Estado$NC))), 2), "%", sep="")
NC_Estado$percent<-percent_Estado</pre>
barplot_NC_Estado_percent<-ggplot(data=NC_Estado)+</pre>
  geom_bar(aes(x=reorder(Estado,-NC),y=NC,fill=Estado),stat = "identity")+
  geom_text(aes(x=Estado,y=NC,label=percent_Estado),hjust=-0.2, size=2.5,angle=90)+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
  labs( x = "State",
        y = "% / Total Loans Granted",
        title ="Loan distribution per State",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
    axis.text.x=element_text(size=8,angle=45,hjust=1),
    axis.text.y=element_text(size=0,angle=45),
    plot.title = element_text(hjust=0.5),
    plot.subtitle=element_text(hjust=0.5),
    legend.position = "none"
    )+ylim(0,625000)
# Por Tipo de Producto
percent TP<-paste(round(100*(NC TP$NC/(sum(NC TP$NC))), 2), "%", sep="")
NC_TP$percent=percent_TP
piechart NC TP<-ggplot(data=NC TP, aes(x="", y=NC, fill=TipoProducto)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_TP), size=3.5, show.legend = F, position = position_stack(vjust = percent_TP)
  guides(fill = guide_legend(title = "Product Type"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
 labs(
   y="".
    x="",
    title="Loan Distribution by Product Type",
    subtitle="2013/01 - 2021/04"
```

```
)+
  theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element text(hjust=0.5),
   axis.text = element_blank(),
   axis.ticks = element_blank(),
   panel.grid = element_blank()
  )
# Por Clasificación de Vivienda
percent_CV<-paste(round(100*(NC_CV$NC/(sum(NC_CV$NC))), 2), "%", sep="")</pre>
NC_CV$percent=percent_CV
piechart_NC_CV<-ggplot(data=NC_CV, aes(x="", y=NC, fill=ClasificacionVivienda)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_CV), size=3.5, show.legend = F, position = position_stack(vjust
  guides(fill = guide_legend(title = "Dwelling Classification"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
  labs(
   y=""
   x="",
   title="Loan Distribution by Dwelling Classification",
   subtitle="2013/01 - 2021/04"
  )+
 theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element_text(hjust=0.5),
   axis.text = element_blank(),
   axis.ticks = element_blank(),
   panel.grid = element_blank()
# Por Linea
percent_L<-paste(round(100*(NC_L$NC/(sum(NC_L$NC))), 2), "%", sep="")</pre>
NC_L$percent=percent_L
piechart_NC_L<-ggplot(data=NC_L, aes(x="", y=NC, fill=Linea)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_L), size=3.5, show.legend = F, position = position_stack(vjust =
  guides(fill = guide_legend(title = "Credit Line"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
  labs(
   y="",
   x="",
   title="Loan Distribution by Credit Line",
   subtitle="2013/01 - 2021/04"
  )+
  theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element_text(hjust=0.5),
   axis.text = element_blank(),
```

```
axis.ticks = element_blank(),
   panel.grid = element_blank()
  )
# Por Ingreso
UMA<-c("1","1-2","2-3","3-4","4-5","5-6","6-7","7-8","8-9","9-10","10-11","11-12","12-13","13-14","14-1
NC_I$UMA<-UMA
barplot_NC_I<-ggplot(data=NC_I)+</pre>
  geom_bar(aes(x=reorder(UMA,-NC),y=NC,fill=UMA),stat = "identity")+
  geom_text(aes(x=UMA,y=NC,label=NC),hjust=-0.2, size=2.5,angle=90)+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(NC_I)))+
  labs( x = "Income (UMA)",
       y = "Number of loans granted",
       title ="Loans distribution by Income (UMA)",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=0,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
  coord cartesian(ylim=c(0,1300000))
# Por Edad
Edad2<-c("20","21-25","26-30","31-35","36-40","41-45","46-50","51-55","56-60","61-65","65+","0tros")
NC_Edad$Edad2<-Edad2
barplot_NC_Edad<-ggplot(data=NC_Edad)+</pre>
  geom_bar(aes(x=Edad2,y=NC,fill=Edad2),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(NC_I)))+
  labs( x = "Age group",
        y = "Number of loans granted",
        title ="Loan distribution by Age",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=6,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
```

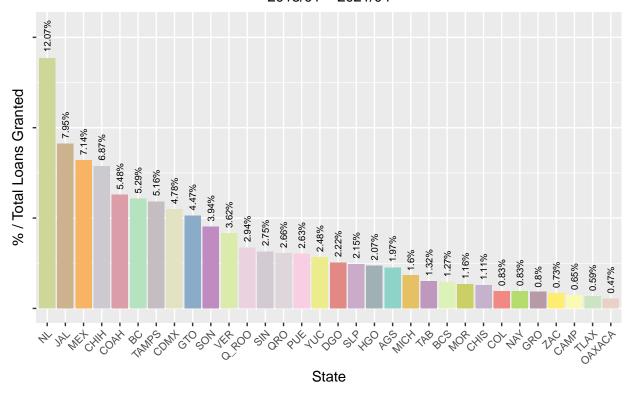
barplot_NC_Estado

Loan distribution per State 2013/01 – 2021/04



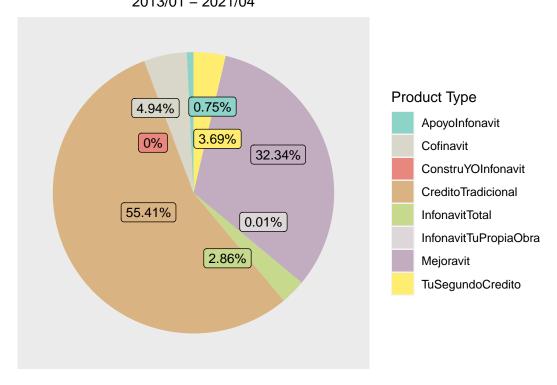
barplot_NC_Estado_percent

Loan distribution per State 2013/01 – 2021/04



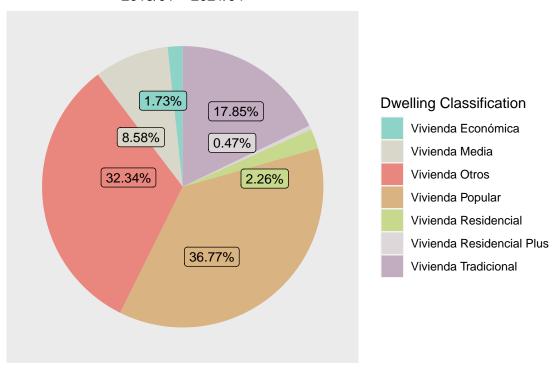
piechart_NC_TP

Loan Distribution by Product Type 2013/01 – 2021/04



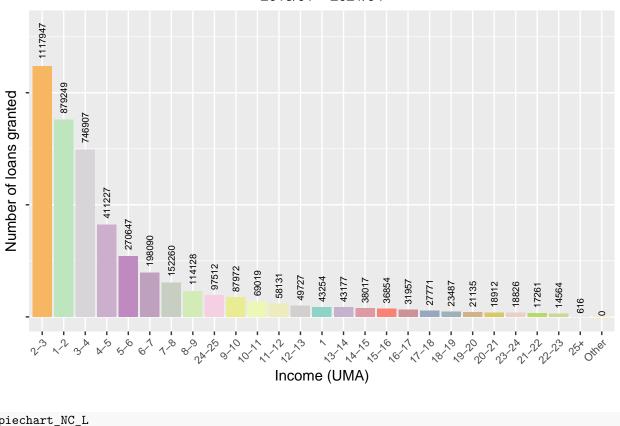
piechart_NC_CV

Loan Distribution by Dwelling Classification 2013/01 – 2021/04



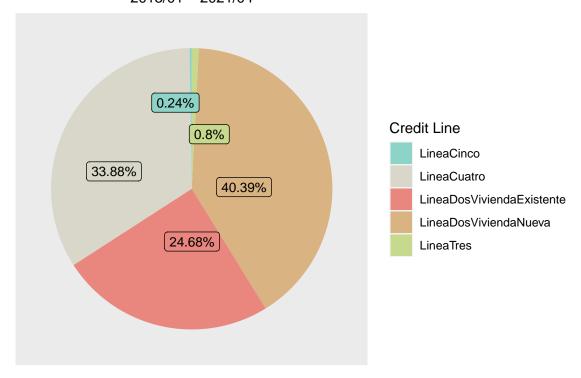
barplot_NC_I

Loans distribution by Income (UMA) 2013/01 - 2021/04



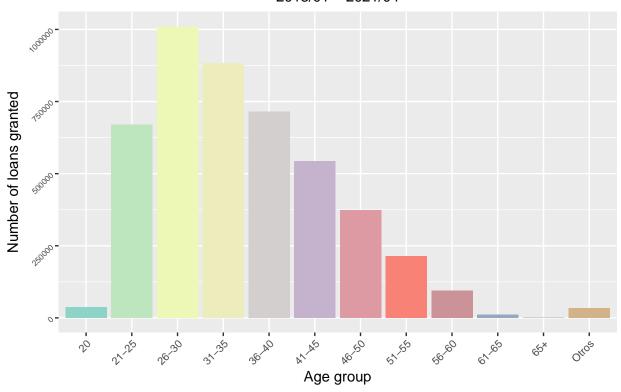
piechart_NC_L

Loan Distribution by Credit Line 2013/01 – 2021/04



barplot_NC_Edad

Loan distribution by Age 2013/01 – 2021/04



1.1.2 Credit Amount (MXN)

```
# Por Estado
barplot_MC_Estado<-ggplot(data=MC_Estado)+</pre>
  geom_bar(aes(x=reorder(Estado,-MC),y=MC,fill=Estado),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
 labs( x = "State",
        y = "Credit amount (MXN)",
        title ="Total Amount of Granted Credit by State",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
    axis.text.x=element_text(size=8,angle=45,hjust=1),
    axis.text.y=element_text(size=8,angle=45),
    plot.title = element_text(hjust=0.5),
    plot.subtitle=element_text(hjust=0.5),
    legend.position = "none"
# Por Estado como porcentaje
percent_Estado2<-paste(round(100*(MC_Estado$MC/(sum(MC_Estado$MC))), 2), "%", sep="")
MC_Estado$percent<-percent_Estado2</pre>
barplot_MC_Estado_percent<-ggplot(data=MC_Estado)+</pre>
```

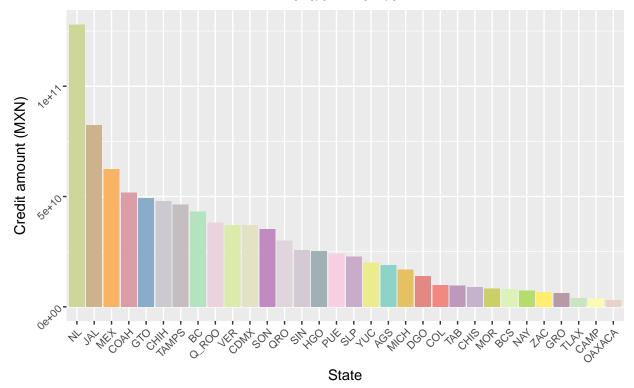
```
geom_bar(aes(x=reorder(Estado,-MC),y=MC,fill=Estado),stat = "identity")+
  geom text(aes(x=Estado,y=MC,label=percent Estado2),hjust=-0.2, size=2.5,angle=90)+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
  labs( x = "State",
        y = "% / Total Amount of Credit",
        title ="Total Amount of Granted Credit by State (%)",
        subtitle = "2013/01 - 2021/04"
  theme(
   axis.text.x=element text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=0,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
   )+ylim(0,14300000000)
# Por Tipo de Producto
percent_TP2<-paste(round(100*(MC_TP$MC/(sum(MC_TP$MC))), 2), "%", sep="")
MC_TP$percent=percent_TP2
piechart_MC_TP<-ggplot(data=MC_TP, aes(x="", y=MC, fill=TipoProducto)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_TP2), size=3, show.legend = F, position = position_stack(vjust =
  guides(fill = guide_legend(title = "Product Type"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
  labs(
   y="".
   x="",
   title="Total Amount of Granted Credit by Product Type",
   subtitle="2013/01 - 2021/04"
  )+
  theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element_text(hjust=0.5),
   axis.text = element_blank(),
   axis.ticks = element_blank(),
   panel.grid = element_blank()
  )
# Por Clasificación de Vivienda
percent_CV2<-paste(round(100*(MC_CV$MC/(sum(MC_CV$MC))), 2), "%", sep="")</pre>
MC CV$percent=percent CV2
piechart_MC_CV<-ggplot(data=MC_CV, aes(x="", y=MC, fill=ClasificacionVivienda)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_CV2), size=3.5, show.legend = F, position = position_stack(vjust
  guides(fill = guide_legend(title = "Dwelling Classification"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
  labs(
   y="",
   x=""
```

```
title="Total Amount of Granted Credit by Dwelling Classification",
    subtitle="2013/01 - 2021/04"
  )+
 theme(
    plot.title = element text(hjust=0.5),
    plot.subtitle = element_text(hjust=0.5),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank()
# Por Linea
percent_L2<-paste(round(100*(MC_L$MC/(sum(MC_L$MC))), 2), "%", sep="")
MC_L$percent=percent_L2
piechart_MC_L<-ggplot(data=MC_L, aes(x="", y=MC, fill=Linea)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_L2), size=3.5, show.legend = F, position = position_stack(vjust = percent_L2)
  guides(fill = guide_legend(title = "Credit Line"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
  labs(
    y="",
    x="",
    title="Total Amount of Granted Credit by Credit Line",
    subtitle="2013/01 - 2021/04"
  )+
 theme(
    plot.title = element_text(hjust=0.5),
    plot.subtitle = element_text(hjust=0.5),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank()
  )
# Por Ingreso
MC I$UMA<-UMA
barplot_MC_I<-ggplot(data=MC_I)+</pre>
  geom_bar(aes(x=reorder(UMA,-MC),y=MC,fill=UMA),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(MC_I)))+
  labs( x = "Income (UMA)",
        y = "Credit Amount (MXN)",
        title ="Total Amount of Granted Credit by Income (UMA)",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
    axis.text.x=element_text(size=8,angle=45,hjust=1),
    axis.text.y=element_text(size=8,angle=45),
    plot.title = element text(hjust=0.5),
    plot.subtitle=element_text(hjust=0.5),
    legend.position = "none"
    )
```

```
# Por Edad
MC_Edad$Edad2<-Edad2
barplot_MC_Edad<-ggplot(data=MC_Edad)+</pre>
  geom_bar(aes(x=Edad2,y=MC,fill=Edad2),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(MC_I)))+
  labs( x = "Age group",
        y = "Credit Amount (MXN)",
        title ="Total Amount of Granted Credit by Age",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
    axis.text.x=element_text(size=8,angle=45,hjust=1),
    axis.text.y=element_text(size=6,angle=45),
    plot.title = element_text(hjust=0.5),
    plot.subtitle=element_text(hjust=0.5),
    legend.position = "none"
```

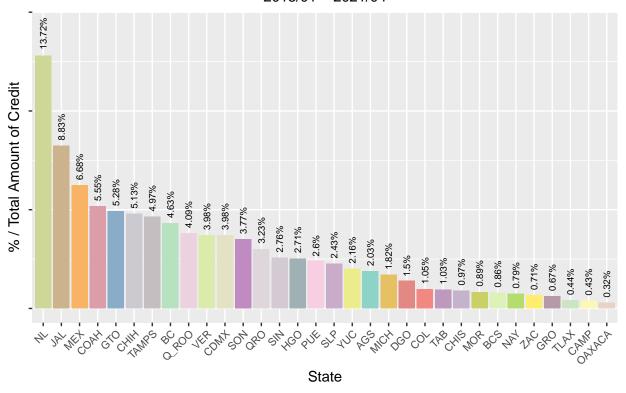
barplot_MC_Estado

Total Amount of Granted Credit by State 2013/01 – 2021/04



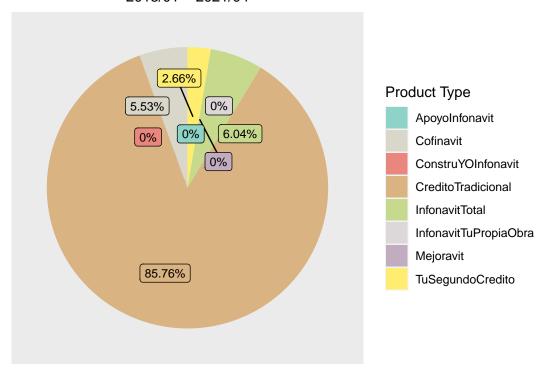
barplot_MC_Estado_percent

Total Amount of Granted Credit by State (%) 2013/01 – 2021/04



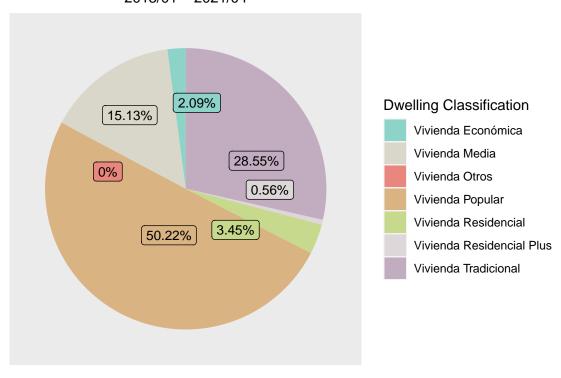
piechart_MC_TP

Total Amount of Granted Credit by Product Type 2013/01 – 2021/04



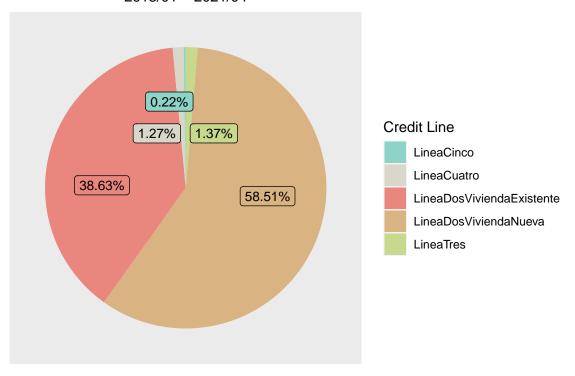
piechart_MC_CV

Total Amount of Granted Credit by Dwelling Classification 2013/01 – 2021/04



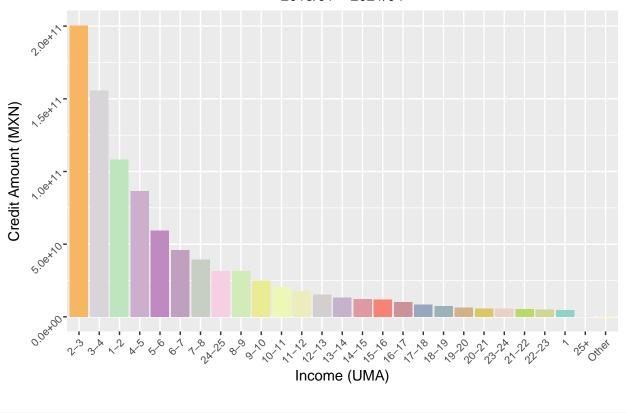
piechart_MC_L

Total Amount of Granted Credit by Credit Line 2013/01 – 2021/04



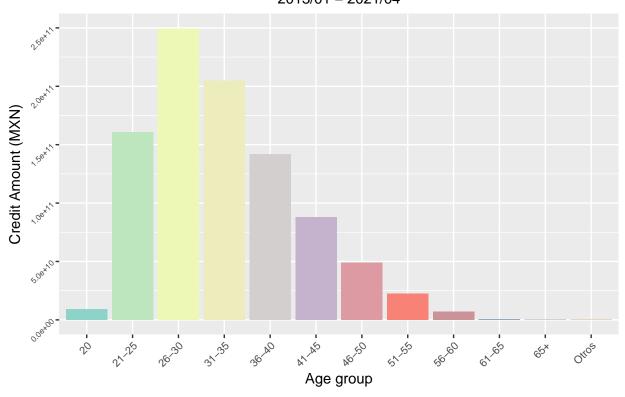
barplot_MC_I

Total Amount of Granted Credit by Income (UMA) 2013/01 – 2021/04



barplot_MC_Edad

Total Amount of Granted Credit by Age 2013/01 – 2021/04



1.1.3 Subsidies

```
# Por Estado
barplot_NS_Estado<-ggplot(data=NS_Estado)+</pre>
  geom_bar(aes(x=reorder(Estado,-NS),y=NS,fill=Estado),stat = "identity")+
 scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
  labs( x = "State",
       y = "Number of Subsidies",
       title ="Number of Granted Subsidies by State",
        subtitle = "2013/01 - 2021/04"
  theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=8,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
# Por Estado como porcentaje
percent_Estado3<-paste(round(100*(NS_Estado$NS)/(sum(NS_Estado$NS))), 2), "%", sep="")
NS_Estado$percent<-percent_Estado3
```

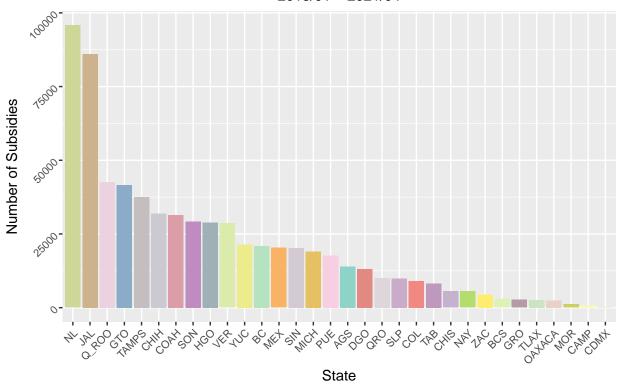
```
barplot_NS_Estado_percent<-ggplot(data=NS_Estado)+</pre>
  geom_bar(aes(x=reorder(Estado,-NS),y=NS,fill=Estado),stat = "identity")+
  geom_text(aes(x=Estado,y=NS,label=percent_Estado3),hjust=-0.2, size=2.5,angle=90)+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
  labs( x = "State",
        y = "% / Total Subsidies Granted",
        title ="Number of Granted Subsidies by State (%)",
        subtitle = "2013/01 - 2021/04"
  theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=0,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
# Por Tipo de Producto
# No Aplica: Todos son crédito tradicional
# Por Clasificación de Vivienda
percent_CV3<-paste(round(100*(NS_CV$NS/(sum(NS_CV$NS))), 2), "%", sep="")</pre>
NS_CV$percent=percent_CV3
piechart_NS_CV<-ggplot(data=NS_CV, aes(x="", y=NS, fill=ClasificacionVivienda)) +
  geom bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_CV3), size=3.5, show.legend = F, position = position_stack(vjust
  guides(fill = guide_legend(title = "Dwelling Classification"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
 labs(
   y="".
   x=""
   title="Distribution of Subsidies by Dwelling Classification",
   subtitle="2013/01 - 2021/04"
  )+
 theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element_text(hjust=0.5),
   axis.text = element_blank(),
   axis.ticks = element_blank(),
   panel.grid = element_blank()
  )
# Por Linea
percent_L3<-paste(round(100*(NS_L$NS/(sum(NS_L$NS))), 2), "%", sep="")
NS_L$percent=percent_L3
piechart_NS_L<-ggplot(data=NS_L, aes(x="", y=NS, fill=Linea)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_L3), size=3.5, show.legend = F, position = position_stack(vjust
  guides(fill = guide_legend(title = "Credit Line"))+
```

```
scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
  labs(
   y="",
   x="",
   title="Distribution of Subsidies by Credit Line",
   subtitle="2013/01 - 2021/04"
  )+
  theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element_text(hjust=0.5),
   axis.text = element_blank(),
   axis.ticks = element_blank(),
   panel.grid = element_blank()
# Por Ingreso
NS_I$UMA<-UMA
barplot_NS_I<-ggplot(data=NS_I)+</pre>
  geom_bar(aes(x=reorder(UMA,-NS),y=NS,fill=UMA),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(NS_I)))+
  labs( x = "Income (UMA)",
       y = "Number of Subsidies",
       title ="Number of Granted Subsidies by Income (UMA)",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=8,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
# Por Edad
NS_Edad$Edad2<-Edad2
barplot_NS_Edad<-ggplot(data=NS_Edad)+
  geom_bar(aes(x=Edad2,y=NS,fill=Edad2),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(NS_I)))+
  labs( x = "Age group",
       y = "Number of Subsidies",
       title ="Number of Granted Subsidies by Age",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
    axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=6,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
```

barplot_NS_Estado

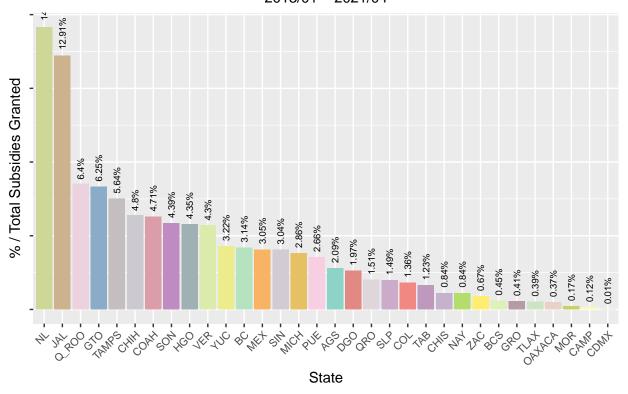
1.1.3.1 Number

Number of Granted Subsidies by State 2013/01 – 2021/04



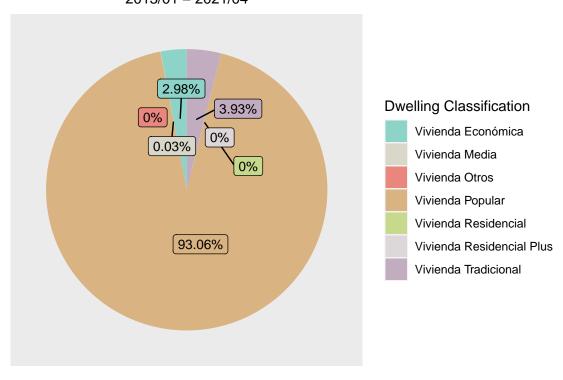
barplot_NS_Estado_percent

Number of Granted Subsidies by State (%) 2013/01 – 2021/04



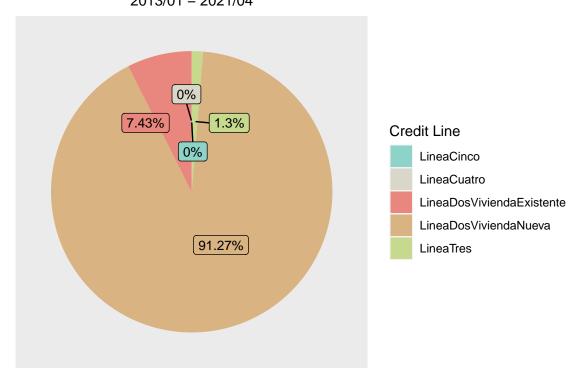
piechart_NS_CV

Distribution of Subsidies by Dwelling Classification 2013/01 – 2021/04



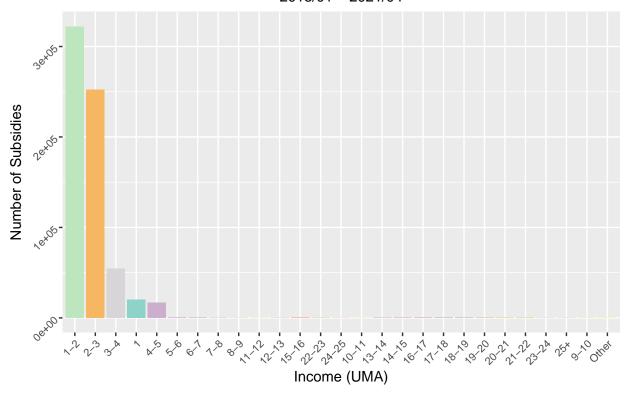
piechart_NS_L

Distribution of Subsidies by Credit Line 2013/01 – 2021/04



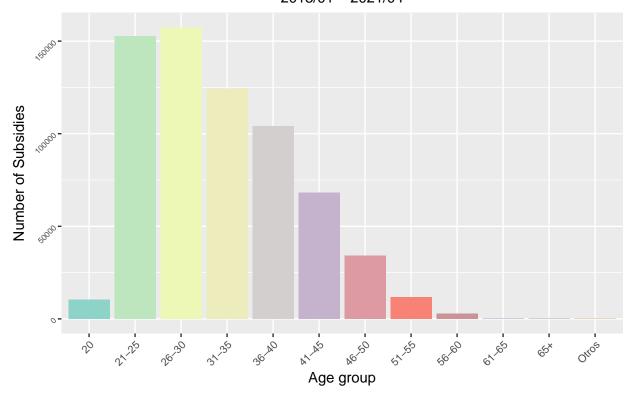
barplot_NS_I

Number of Granted Subsidies by Income (UMA) 2013/01 – 2021/04



barplot_NS_Edad

Number of Granted Subsidies by Age 2013/01 – 2021/04



```
# Por Estado
barplot_MS_Estado<-ggplot(data=MS_Estado)+</pre>
  geom_bar(aes(x=reorder(Estado,-MS),y=MS,fill=Estado),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
 labs( x = "State",
       y = "Amount (MXN)",
       title ="Amount of Granted Subsidies by State",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=8,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
# Por Estado como porcentaje
percent_Estado4<-paste(round(100*(MS_Estado$MS/(sum(MS_Estado$MS))), 2), "%", sep="")
MS_Estado$percent<-percent_Estado4
barplot_MS_Estado_percent<-ggplot(data=MS_Estado)+
  geom_bar(aes(x=reorder(Estado,-MS),y=MS,fill=Estado),stat = "identity")+
```

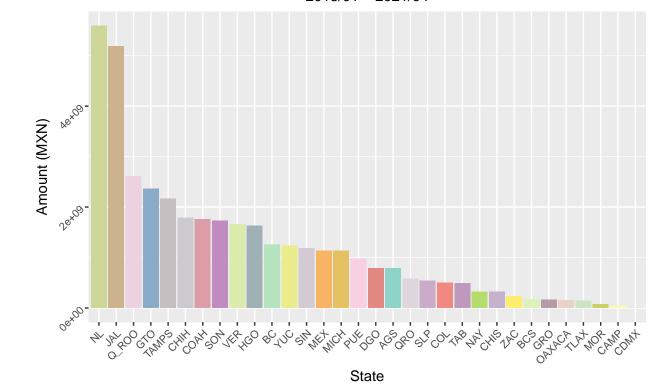
```
geom_text(aes(x=Estado,y=MS,label=percent_Estado4),hjust=-0.2, size=2.5,angle=90)+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(32))+
  labs( x = "State",
       v = "% / Total Amount",
        title ="Amount of Granted Subsidies by State (%)",
        subtitle = "2013/01 - 2021/04"
        )+
  theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=0,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
# Por Tipo de Producto
# No Aplica: Todos son crédito tradicional
# Por Clasificación de Vivienda
percent_CV4<-paste(round(100*(MS_CV$MS/(sum(MS_CV$MS))), 2), "%", sep="")
MS_CV$percent=percent_CV4
piechart_MS_CV<-ggplot(data=MS_CV, aes(x="", y=MS, fill=ClasificacionVivienda)) +
  geom_bar(stat="identity", width=1) +
  coord polar("y", start=0)+
  geom_label_repel(aes(label = percent_CV4), size=3.5, show.legend = F, position = position_stack(vjust
  guides(fill = guide legend(title = "Dwelling Classification"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
  labs(
   y="",
   x="",
   title="Amount of Subsidies by Dwelling Classification",
   subtitle="2013/01 - 2021/04"
  )+
 theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element_text(hjust=0.5),
   axis.text = element_blank(),
   axis.ticks = element_blank(),
   panel.grid = element_blank()
  )
# Por Linea
percent L4<-paste(round(100*(MS L$MS/(sum(MS L$MS))), 2), "%", sep="")
MS_L$percent=percent_L4
piechart_MS_L<-ggplot(data=MS_L, aes(x="", y=MS, fill=Linea)) +</pre>
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  geom_label_repel(aes(label = percent_L4), size=3.5, show.legend = F, position = position_stack(vjust
  guides(fill = guide_legend(title = "Credit Line"))+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(8))+
```

```
x="",
   title="Amount of Subsidies by Credit Line",
   subtitle="2013/01 - 2021/04"
  )+
 theme(
   plot.title = element_text(hjust=0.5),
   plot.subtitle = element_text(hjust=0.5),
   axis.text = element_blank(),
   axis.ticks = element_blank(),
   panel.grid = element_blank()
  )
# Por Ingreso
MS_I$UMA<-UMA
barplot_MS_I<-ggplot(data=MS_I)+</pre>
  geom_bar(aes(x=reorder(UMA,-MS),y=MS,fill=UMA),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(MS_I)))+
  labs( x = "Income (UMA)",
        y = "Amount (MXN)",
       title = "Amount of Granted Subsidies by Income",
        subtitle = "2013/01 - 2021/04"
  theme(
   axis.text.x=element text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=8,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
   )
# Por Edad
MS_Edad$Edad2<-Edad2
barplot_MS_Edad<-ggplot(data=MS_Edad)+</pre>
  geom_bar(aes(x=Edad2,y=MS,fill=Edad2),stat = "identity")+
  scale_fill_manual(values = colorRampPalette(brewer.pal(12, "Set3"))(nrow(MS_I)))+
 labs( x = "Age group",
        y = "Amount (MXN)",
       title ="Amount of Granted Subsidies by Age",
       subtitle = "2013/01 - 2021/04"
       )+
 theme(
   axis.text.x=element_text(size=8,angle=45,hjust=1),
   axis.text.y=element_text(size=6,angle=45),
   plot.title = element_text(hjust=0.5),
   plot.subtitle=element_text(hjust=0.5),
   legend.position = "none"
```

```
barplot_MS_Estado
```

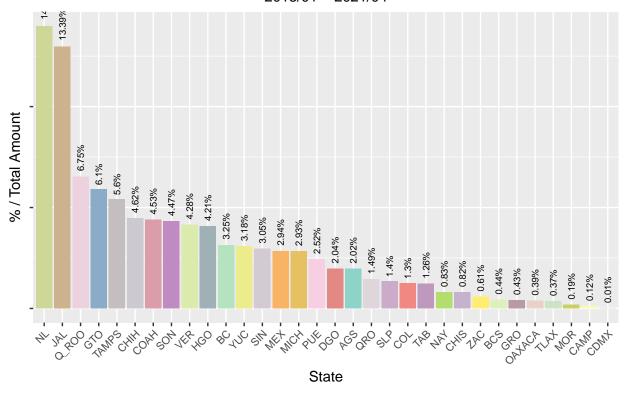
1.1.3.2 Amount (\$MXN)

Amount of Granted Subsidies by State 2013/01 – 2021/04



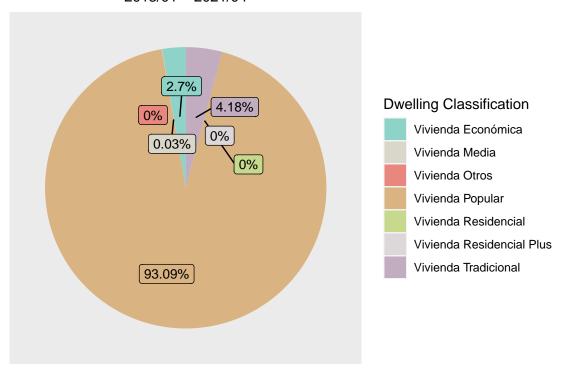
barplot_MS_Estado_percent

Amount of Granted Subsidies by State (%) 2013/01 – 2021/04



piechart_MS_CV

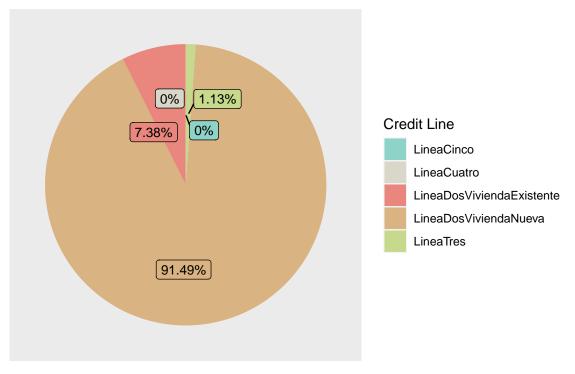
Amount of Subsidies by Dwelling Classification 2013/01 – 2021/04



piechart_MS_L

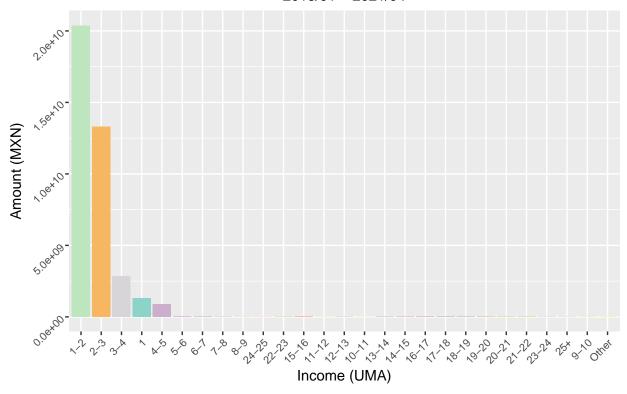
Amount of Subsidies by Credit Line

2013/01 - 2021/04



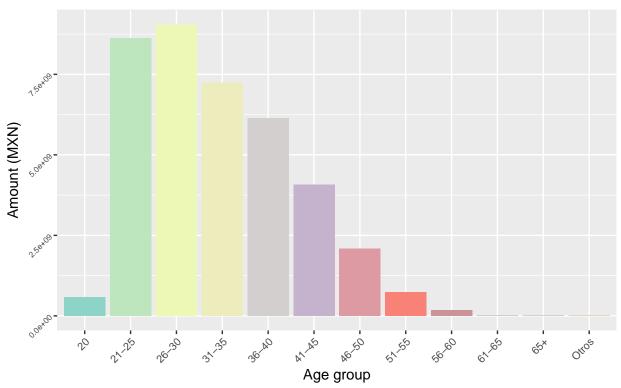
barplot_MS_I

Amount of Granted Subsidies by Income 2013/01 – 2021/04



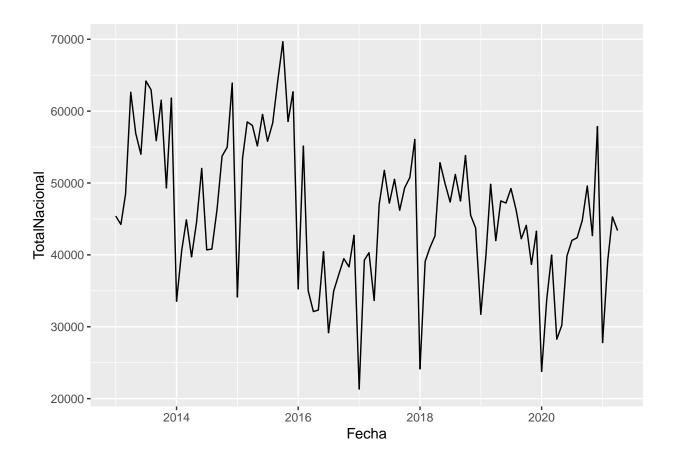
barplot_MS_Edad

Amount of Granted Subsidies by Age 2013/01 – 2021/04



Solo para ver seasonality en el número de créditos otorgados por mes:

```
timeseries_NCTotal<-ggplot(data=NCTotal)+
  geom_line(aes(x=Fecha,y=TotalNacional))
timeseries_NCTotal</pre>
```



1.2 Data Reported in Sistema de Información Infonavit: Complete Portfolio Analysis

Leaving behind the database of "otorgamiento de crédito" we now consider the information reported in the section "Estadísticas", which gives portfolio operating indicators of Infonavit's **complete** portfolio. The database decomposes Infonavit's portfolio by: past due portfolio (vencida), extension portfolio (prórrroga) and current portfolio (vigente).

Note that, before, we were analyzing credits originated from January 2013 to April 2021, but this new database includes mortgages originated **before** 2013, considering the **complete portfolio** as of 30th of April, 2021.

```
# Complete Portfolio (CP) composition

CP_cuentas<-as.data.frame(read_excel("DATA/CarteraVencida.xlsx", sheet = 1))

CP_saldos<-as.data.frame(read_excel("DATA/CarteraVencida.xlsx", sheet = 2))

# Panel con cuatro gráficas, una gráfica por año: Barras por estado con el total

# de cartera en número de cuentas. Cada barra apilada por vigente + prórroga + vencida

# PENDIENTE

# Panel con cuatro gráficas, una gráfica por año: Barras por estado con el total
```

```
# de cartera en saldo (mdp MXN). Cada barra apilada por vigente + prórroga + vencida
# PENDIENTE
# Gráfica con la sheet 2 del excel, pero sos*
# PENDIENTE
```

2 Sample data

Now, we consider our main database. This is the data from where we intend to develop our credit risk model; it describes credits originated from January 2013 to April 2021 *individually*, characterizing each beneficiary by different variables like age, gender, type of dwelling, etc.

In theory, this data base should be exactly the same as the data analyzed at the beginning, reported in "Sistema de Información Infonavit" for 2013/01-2021/04. However, we do not expect this database to be distributed exactly the same as the **complete portfolio** information. We intend to manipulate this sample to best resemble the complete portfolio, so that the credit risk model can more accurately represent reality.

2.1 Data Analysis

```
sample<-readRDS(file="DATA/MUESTRA.Rds")
head(sample)</pre>
```

```
AÑO MES_ACUMULADO CLAVE_ENTIDAD CLAVE_MUNICIPIO
                                                               MODALIDAD
## 1 2013
                       1
                                     1
                                                      1 Viviendas nuevas
## 2 2013
                       1
                                     1
                                                      1 Viviendas nuevas
## 3 2013
                       1
                                     1
                                                      1 Viviendas nuevas
## 4 2013
                                     1
                                                      1 Viviendas nuevas
## 5 2013
                       1
                                     1
                                                      1 Viviendas nuevas
## 6 2013
                                     1
                                                      5 Viviendas nuevas
                                              TIPO GENERO
##
            DESTINO
                                                                 EDAD
                                                                           TNGRESO
## 1 Vivienda nueva Cofinanciamientos y subsidios Hombre 29 o menos 2.61 a 4.00
## 2 Vivienda nueva Cofinanciamientos y subsidios Hombre 29 o menos 4.01 a 6.00
## 3 Vivienda nueva Cofinanciamientos y subsidios Hombre 29 o menos 6.01 a 9.00
## 4 Vivienda nueva Cofinanciamientos y subsidios Hombre 29 o menos
## 5 Vivienda nueva Cofinanciamientos y subsidios Hombre 29 o menos
                                                                        M·s de 12
##
  6 Vivienda nueva Cofinanciamientos y subsidios Hombre 29 o menos 6.01 a 9.00
##
     VALOR_VIVIENDA
                        MONTO CLAVE
## 1
        Tradicional
                     283223.1
## 2
            Popular 1365780.9
                                 1.1
## 3
                     803068.0
        Tradicional
                                 1.1
## 4
        Tradicional
                     671815.4
                                 1.1
## 5
              Media 1650390.9
                                 1.1
## 6
          EconÛmica 214819.1
                                 1.5
```

```
tail(sample)
```

```
## AÑO MES_ACUMULADO CLAVE_ENTIDAD CLAVE_MUNICIPIO MODALIDAD ## 1397063 2021 1 28 27 Viviendas usadas ## 1397064 2021 1 3 1 Viviendas usadas
```

```
## 1397065 2021
                                          14
                                                          97 Viviendas usadas
## 1397066 2021
                            2
                                          28
                                                           35 Viviendas usadas
                                                          68 Viviendas usadas
## 1397067 2021
                            1
                                          30
## 1397068 2021
                                          14
                                                           44 Viviendas usadas
                  DESTINO
                                         TIPO GENERO
                                                           EDAD
                                                                     INGRESO
## 1397063 Vivienda usada Credito individual Hombre 29 o menos 2.6 o menos
## 1397064 Vivienda usada Credito individual Hombre 29 o menos 2.6 o menos
## 1397065 Vivienda usada Credito individual Mujer 29 o menos 2.6 o menos
## 1397066 Vivienda usada Credito individual
                                               Mujer 29 o menos 2.6 o menos
## 1397067 Vivienda usada Credito individual
                                               Mujer 29 o menos 2.6 o menos
## 1397068 Vivienda usada Credito individual
                                               Mujer 29 o menos 2.6 o menos
           VALOR_VIVIENDA
                             MONTO CLAVE
                EconÛmica 312690.0 28.27
## 1397063
## 1397064
                EconÛmica 316308.3
## 1397065
                EconÛmica 319551.4 14.97
## 1397066
                EconÛmica 322922.0 28.35
## 1397067
                EconÛmica 275690.8 30.68
                EconÛmica 300287.6 14.44
## 1397068
str(sample)
## 'data.frame':
                    1397068 obs. of 13 variables:
                     : Factor w/ 9 levels "2013","2014",...: 1 1 1 1 1 1 1 1 1 1 ...
    $ AÑO
                     : Factor w/ 12 levels "1", "2", "3", "4", ...: 1 1 1 1 1 1 1 1 1 1 ...
    $ MES ACUMULADO
                     : Factor w/ 32 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
##
    $ CLAVE ENTIDAD
    $ CLAVE_MUNICIPIO: Factor w/ 316 levels "0","1","2","3",..: 2 2 2 2 2 6 7 12 12 ...
   $ MODALIDAD
                     : Factor w/ 4 levels "Mejoramientos",..: 3 3 3 3 3 3 3 3 3 ...
                     : Factor w/ 6 levels "Autoproducci\hat{\mathbb{U}}n",...: 5 5 5 5 5 5 5 5 5 5 ...
   $ DESTINO
##
    $ TIPO
                     : Factor w/ 2 levels "Cofinanciamientos y subsidios",..: 1 1 1 1 1 1 1 1 1 1 ...
##
   $ GENERO
                     : Factor w/ 3 levels "Hombre", "Mujer", ...: 1 1 1 1 1 1 1 1 1 1 ...
##
  $ EDAD
                     : Factor w/ 4 levels "29 o menos", "30 a 59", ...: 1 1 1 1 1 1 1 1 1 1 ...
                     : Factor w/ 7 levels "2.6 o menos",..: 2 3 4 6 6 4 5 3 3 5 ...
##
    $ INGRESO
    $ VALOR VIVIENDA: Factor w/7 levels "EconÔmica", "Media", ...: 7 4 7 7 2 1 7 4 7 7 ...
##
    $ MONTO
                     : num 283223 1365781 803068 671815 1650391 ...
                     : Factor w/ 1911 levels "1.0","1.1","1.10",...: 2 2 2 2 2 8 9 4 4 4 ...
##
    $ CLAVE
summary(sample)
                                                         CLAVE MUNICIPIO
##
         AÑO
                     MES_ACUMULADO
                                       CLAVE_ENTIDAD
##
   2019
           :188218
                     12
                            :133591
                                              :149449
                                                                : 47494
    2018
           :186917
                     10
                             :125325
                                       19
                                              :115360
                                                                  46046
                                                         5
    2015
##
                                                         2
                                                                   44994
           :186073
                     8
                             :119602
                                       14
                                              : 99518
##
    2017
           :179491
                     2
                             :119336
                                       30
                                              : 81642
                                                                   41861
                                                         1
##
    2020
           :179336
                             :118882
                                              : 74938
                                                                   37877
    2016
##
                                              : 64865
                                                                  36384
           :156016
                             :118240
                                       11
                                                         18
##
    (Other):321017
                     (Other):662092
                                       (Other):811296
                                                         (Other):1142412
                                                         DESTINO
##
               MODALIDAD
##
  Mejoramientos
                    :361602
                               AutoproducciÛn
                               Con disponibilidad de terreno: 16626
## Otros programas: 9766
## Viviendas nuevas:496191
                               Mejoramientos
                                                             :361602
##
  Viviendas usadas:529509
                               Pago de pasivos
                                                             : 9766
##
                               Vivienda nueva
                                                             :479560
##
                               Vivienda usada
                                                             :529509
```

```
##
                                                     GENERO
##
                               TIPO
    Cofinanciamientos y subsidios:500133
##
                                           Hombre
                                                         :858539
    Credito individual
                                 :896935
##
                                           Mujer
                                                         :531362
##
                                           No disponible: 7167
##
##
##
##
##
               EDAD
                                    INGRESO
                                                            VALOR_VIVIENDA
##
    29 o menos
                 :417938
                           2.6 o menos
                                        :318690
                                                  EconÛmica
                                                                 : 65335
                 :945999
                           2.61 a 4.00
                                        :270913
##
    30 a 59
                                                  Media
                                                                   :197201
                 : 12652
    60 o m·s
                           4.01 a 6.00 :244657
                                                  No disponible
##
                                                                  :324806
##
    No disponible: 20479
                           6.01 a 9.00 :222882
                                                  Popular
                                                                   :376124
                           9.01 a 12.00 :134704
##
                                                  Residencial
                                                                   : 48282
##
                           M·s de 12
                                        :205114
                                                  Residencial plus: 9013
##
                           No disponible:
                                            108
                                                  Tradicional
                                                                   :376307
                            CLAVE
##
        MONTO
##
          : -424633
                        14.120 : 16668
   Min.
    1st Qu.:
               58243
                        2.4
                                  16586
##
   Median :
               352699
                        11.20 :
                                  16072
    Mean
               763071
                        8.19
                                  15985
               740000
    3rd Qu.:
                        19.6
                               : 15464
##
    Max. :180487760
                        22.14 : 15023
##
                        (Other):1301270
# GRÁFICAS DE GABRIEL* Pegar su code*
```

2.2 Adjust Sample to Match Population's Characteristics

```
# Berny
```

2.3 Asign Variable Status

```
# Luly
```

- 3 Credit Risk Model
- 3.1 Logistic Regression
- 3.2 Evaluation
- 3.3 Regional Analysis & Clusters
- 4 Building a Cross-Subsidy Scheme
- 4.1 Asigning an Interest Rate
- ${\bf 5}\quad {\bf Comparative\ Statics\ \&\ Tests}$