

# Introducción a los modelos mixtos (SIA 3011003)

Profesor Juan Carlos Salazar-Uribe  
jcsalaza@unal.edu.co



# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

"The last few decades have seen the rise of formal theories of statistics,"legitimizing" variation by confining it by assumption to random sampling, often assumed to involve tightly specified distributions, and restoring the appearance of security by emphasizing narrowly optimized techniques and claiming to make statements with "known" probabilities of error. While some of the influences of statistical theory on data analysis have been helpful, others have not. Exposure<sup>1</sup>, the effective laying open of the data to display the unanticipated, is to us a major portion of data analysis. Formal statistics has given almost no guidance to exposure; indeed, it is not clear how the informality and flexibility appropriate to the exploratory character of exposure can be fitted into any of the structures of formal statistics so far proposed."

The Future of Data Analysis, John W. Tukey 1962 <sup>2</sup>.

---

<sup>1</sup>"The term "exposure" can be applied to the primary explanatory variable of interest and to other variables that may be associated with the outcome, such as confounders or effect modifiers, which also must be addressed in the analysis of the primary outcome."

<sup>2</sup><https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-33/issue-1/The-Future-of-Data-Analysis/10.1214/aoms/1177704711.full>

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

En la práctica a menudo se recolectan datos desequilibrados o desbalanceados en el sentido de que hay un número desigual de mediciones por sujeto y mediciones no tomadas en puntos de tiempo fijos (no equiespaciadas). Por lo tanto, las técnicas de regresión multivariadas a menudo no son aplicables. Sin embargo, los MLM pueden manejar estas situaciones. También, a menudo, los perfiles longitudinales específicos del sujeto se pueden aproximar bien mediante funciones de regresión. Esta idea se explorará más en detalle a medida que transcurra en curso.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

En estadística, el análisis exploratorio de datos (EDA<sup>3</sup>) es un enfoque de análisis de datos para resumir sus características principales, a menudo utilizando gráficos estadísticos y otros métodos de visualización de datos. Se puede usar o no un modelo estadístico, pero principalmente EDA es para ver lo que los datos pueden decirnos más allá del modelamiento formal o la implementación de pruebas de hipótesis. John Tukey ha promovido el análisis exploratorio de datos desde 1970 para alentar a los estadísticos a explorar los datos y posiblemente formular hipótesis que podrían conducir a nuevos experimentos y recopilación de datos.

---

<sup>3</sup>Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley.

Tukey definió el análisis de datos en 1961 como: “Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.”<sup>4</sup>

---

<sup>4</sup>Tukey, John W. (1977). Exploratory Data Analysis. Pearson. ISBN 978-0201076165.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Los objetivos de EDA son:

- Sugerir hipótesis sobre las causas de los fenómenos observados.
- Evaluar los supuestos en los que se basará la inferencia estadística.
- Apoyar la selección de herramientas y técnicas estadísticas apropiadas.
- Proporcionar una base para una mayor recopilación de datos a través de encuestas o experimentos.

Muchas técnicas EDA se han adoptado en la [minería de datos](#). También se les enseña a jóvenes estudiantes como una forma de introducirlos al pensamiento estadístico.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Como parte de las **buenas prácticas estadísticas**, siempre se recomienda realizar un completo análisis descriptivo de los datos antes de implementar cualquier modelo. De esta forma se pueden anticipar posibles problemas con la calidad de los datos, así como observar patrones que permitan dilucidar soluciones o respuestas a las preguntas de investigación o de interés. Este análisis descriptivo, usualmente se compone de medidas numéricas de resumen y de gráficos (**visualización**).

¿QUÉ ES LA VISUALIZACIÓN DE DATOS?. La visualización de datos, es el proceso de crear representaciones gráficas de información. Este proceso ayuda al presentador a comunicar datos de una manera que le facilite, a quien recibe esta información, su interpretación, la generación de conclusiones y le oriente en el proceso de toma de decisiones.



# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Entre las medidas numéricas de resumen más comunes para variables de tipo continuo, se encuentran el promedio, la mediana, la moda, el mínimo, el máximo, la desviación estándar, los cuartiles, el rango, entre otros, que son flexibles y viables de calcular aún en presencia de datos longitudinales. Las variables categóricas se acostumbra describirlas con frecuencias y porcentajes así como con tablas de contingencia. Es muy común y recomendable estudiar variables continuas segmentadas o estratificadas de acuerdo a los niveles de variables categóricas de interés científico directo.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Por su parte, la visualización ofrece una amplia gama de gráficos versátiles e informativos que permiten miradas de los datos desde distintas perspectivas lo cual potencia el análisis de datos y confirma el papel de la estadística como una poderosa herramienta, imprescindible, en el proceso de generación de conocimiento a partir de datos, competencia esta muy apetecida y valorada hoy en día en la era del internet. El siguiente recurso de R es altamente recomendable para este fin: <https://www.r-graph-gallery.com/>.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Estadísticos de resumen.** Se usarán los datos de Spruce Sitka (abetos) para ilustrar cómo en R se pueden obtener algunos de los estadísticos de resumen más comunes cuando se trabaja con datos longitudinales.

```
#Ejemplos clase 1 INTROMLM
library(lattice)
library(ggplot2)
library(gridExtra)
library(Rcpp)
library(MASS)
library(lattice)

data_wide<-read.csv(file="Sitka_Spruce_ancho.csv",header=T,sep=',',dec='.')
sapply(data_wide, class)
print("Frecuencias por grupo")
table(data_wide$Group)
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen.

```
##          ID          Group      y152      y174      y201      y227
##  "integer" "character"  "numeric"  "numeric"  "numeric"  "numeric"
```

```
## [1] "Frecuencias por grupo"
```

```
##
## Control  Ozone
##      12      12
```

## Estadísticos de resumen.

```
#Ejemplos clase 1 INTROMLM
```

```
library(lattice)  
library(ggplot2)  
library(gridExtra)  
library(Rcpp)  
library(MASS)  
library(lattice)
```

```
data_wide<-read.csv(file="Sitka_Spruce_ancho.csv",header=T,sep=',',dec='.')  
sapply(data_wide, summary)
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen.

```
## $ID
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   6.75   12.50   12.50   18.25   24.00
##
## $Group
##      Length      Class    Mode
##      24 character character
##
## $y152
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.790  4.093  4.390  4.381  4.625   5.460
##
## $y174
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.100  4.490  4.860  4.789  5.090   5.790
##
## $y201
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.300  4.915  5.140  5.155  5.405   6.120
##
## $y227
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.380  5.220  5.495  5.495  5.933   6.410
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen. Correlaciones

```
#Ejemplos clase 1 INTROMLM
library(lattice)
library(ggplot2)
library(gridExtra)
library(Rcpp)
library(MASS)
library(lattice)

data_wide<-read.csv(file="Sitka_Spruce_ancha.csv",header=T,sep=',',dec='.')
#head(data_wide)
print("Correlaciones entre las mediciones en el tiempo")
print("Pearson")
round(cor(data_wide[,3:6],method="pearson"),3)
print("Spearman")
round(cor(data_wide[,3:6],method="spearman"),3)
print("Kendall")
round(cor(data_wide[,3:6],method="kendall"),3)
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Estadísticos de resumen.** En las tres matrices de correlación se observa, que entre más separadas las observaciones, la correlación disminuye. Es decir, la correlación a los tiempos  $t_i$  y  $t_j$  depende de  $|t_j - t_i|$  (Estacionaridad débil)

```
## [1] "Correlaciones entre las mediciones en el tiempo"
```

```
## [1] "Pearson"
```

```
##      y152  y174  y201  y227
## y152 1.000 0.981 0.961 0.919
## y174 0.981 1.000 0.976 0.948
## y201 0.961 0.976 1.000 0.966
## y227 0.919 0.948 0.966 1.000
```

```
## [1] "Spearman"
```

```
##      y152  y174  y201  y227
## y152 1.000 0.965 0.918 0.881
## y174 0.965 1.000 0.964 0.928
## y201 0.918 0.964 1.000 0.946
## y227 0.881 0.928 0.946 1.000
```

```
## [1] "Kendall"
```

```
##      y152  y174  y201  y227
## y152 1.000 0.868 0.780 0.746
## y174 0.868 1.000 0.873 0.788
## y201 0.780 0.873 1.000 0.838
## y227 0.746 0.788 0.838 1.000
```



# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen.

```
## [1] "Estadísticos de resumen por grupo"
```

```
## data$Group: Control
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	2.790	4.303	4.780	4.668	5.117	6.120

```
## -----
```

```
## data$Group: Ozone
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	4.000	4.893	5.185	5.242	5.683	6.410

```
## [1] "Desviaciones estándar por grupo"
```

```
## data$Group: Control
```

```
## [1] 0.7384724
```

```
## -----
```

```
## data$Group: Ozone
```

```
## [1] 0.5864797
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen.

```
## [1] "Estadísticos de resumen de acuerdo al tiempo"
```

```
## data$time1: 152
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	2.790	4.093	4.390	4.381	4.625	5.460

```
## -----
```

```
## data$time1: 174
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3.100	4.490	4.860	4.789	5.090	5.790

```
## -----
```

```
## data$time1: 201
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3.300	4.915	5.140	5.155	5.405	6.120

```
## -----
```

```
## data$time1: 227
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3.380	5.220	5.495	5.495	5.933	6.410

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen.

```
## [1] "Desviaciones estándar de acuerdo al tiempo"
```

```
## data$time1: 152
```

```
## [1] 0.5781553
```

```
## -----
```

```
## data$time1: 174
```

```
## [1] 0.5803396
```

```
## -----
```

```
## data$time1: 201
```

```
## [1] 0.59268
```

```
## -----
```

```
## data$time1: 227
```

```
## [1] 0.6489054
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen.

*#Ejemplos clase 3 INTROMLM*

```
library(questionr)
print("Tabla de doble entrada de frecuencias Grupo x Tiempo. Abetos")
wtd.table(x = data$time1, y = data$Group)
print("Tabla de doble entrada con pesos Grupo x Tiempo. Abetos")
wtd.table(x = data$time1, y = data$Group, weights = data$log_growth)
sum(data_wide[data_wide$Group=="Control",3])
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Estadísticos de resumen.

```
## [1] "Tabla de doble entrada de frecuencias Grupo x Tiempo. Abetos"
```

```
##      Control Ozone
## 152      12      12
## 174      12      12
## 201      12      12
## 227      12      12
```

```
## [1] "Tabla de doble entrada con pesos Grupo x Tiempo. Abetos"
```

```
##      Control Ozone
## 152    49.02 56.13
## 174    53.81 61.12
## 201    58.81 64.91
## 227    62.41 69.47
```

```
## [1] 49.02
```

Cada entrada de esta última tabla corresponde a **la suma de los valores observados** respecto al cruce de los niveles de las respectivas categorías

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Estadísticos de resumen. Tabla de doble entrada de promedios  
Grupo x Tiempo. Abetos

```
#Ejemplos clase 3 INTROMLM
library(questionr)
print("Tabla de doble entrada de promedios Grupo x Tiempo. Abetos")
a<-wtd.table(x = data$time1, y = data$Group)
b<-wtd.table(x = data$time1, y = data$Group, weights = data$log_growth)
round(b/a,2)
```

## Estadísticos de resumen. Tabla de doble entrada de promedios Grupo x Tiempo. Abetos

```
## [1] "Tabla de doble entrada de promedios Grupo x Tiempo. Abetos"
```

```
##      Control Ozone
## 152      4.08  4.68
## 174      4.48  5.09
## 201      4.90  5.41
## 227      5.20  5.79
```

**Visualización de datos longitudinales.** Primero se describirán **algunos** de los gráficos más comunes en visualización y luego se ilustrará como obtenerlos usando software estándar tal como R. **Es necesario insistir en la importancia de graficar los datos antes de proceder a realizar cualquier implementación de algún modelo estadístico.**



# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Gráfico de valores missing.** Este gráfico es muy útil, ya que permite identificar qué variable en una base de datos tiene o no datos perdidos y en qué proporción.

```
library(Amelia)
data_wide<-read.csv(file="Sitka_Spruce_ancho.csv",
                    header=T,sep=',',dec='.')
missmap(data_wide)
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de valores missing.

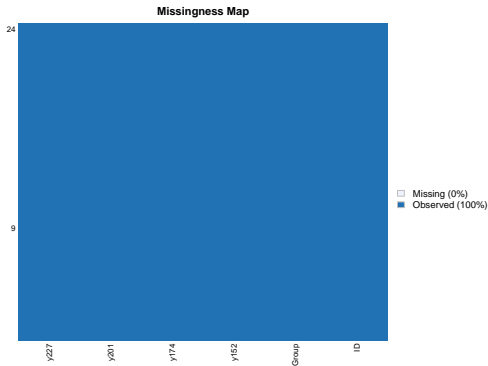


Figura 1: Gráfico de valores perdidos. Abetos

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Gráfico de valores missing.** Si la base de datos con que se cuenta, tiene valores missing en alguna de sus variables, el missmap luciría como:

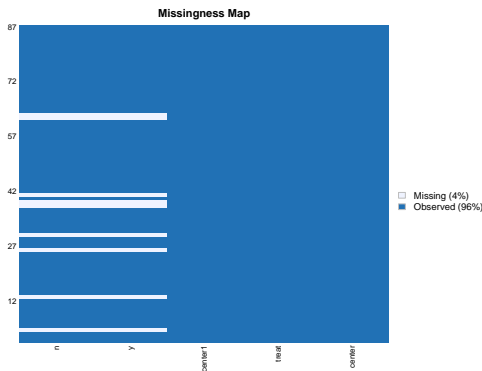


Figura 2: Gráfico de valores perdidos. Colfeet dataset

**Gráfico de barras apilado.** Es un gráfico que permite representar valores de una variable de acuerdo a dos variables categóricas. Generalmente se usa para conteos, pero se puede adaptar para otros estadísticos tales como la media o la mediana.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Gráfico de barras apilado

```
library(ggplot2)
library(ggpubr)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$i.ID
data$Time<-as.factor(data$time1)

ggbarplot(data, x = "Group", y = "log_growth", fill = "Time", label = TRUE,
  add = "mean", lab.pos = c("in"),
  lab.col = "red", lab.nb.digits=2,
  palette = "Paired", width=0.5)
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Gráfico de barras apilado

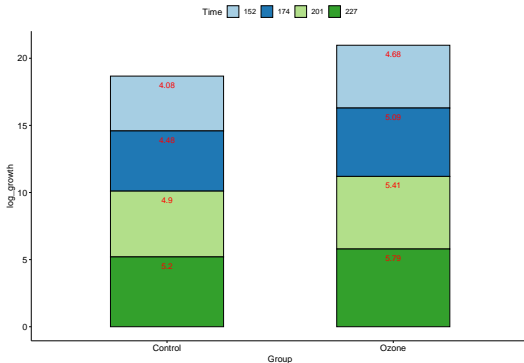


Figura 3: Gráfico de barras apilado respecto al Grupo y Tiempo. Abetos

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Gráfico de barras apilado

```
library(ggplot2)
library(ggpubr)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$i.ID
data$Time<-as.factor(data$time1)

ggbarplot(data, x = "Time", y = "log_growth", fill = "Group", label = TRUE,
  add = "mean", lab.pos = c("in"),
  lab.col = "red", lab.nb.digits=2,
  palette = "Paired", width=0.5)
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Gráfico de barras apilado

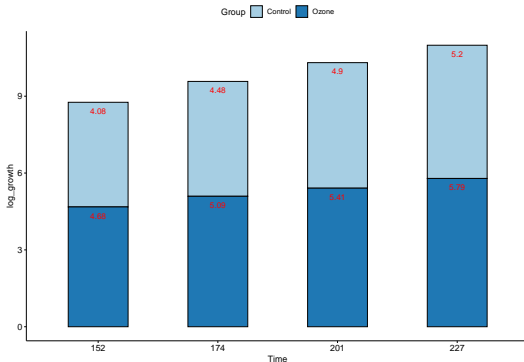


Figura 4: Gráfico de barras apilado respecto al Tiempo y Grupo. Abetos



# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Gráfico de barras apilado con suma acumulada

```
library(questionr)
library(ggplot2)
library(gridExtra)
library(ggpubr)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$i..ID
data$Time<-as.factor(data$time1)
#Cumulative sum of log-growth by time according to Group
cumsum<-data.frame(wtd.table(x = data$time1, y = data$Group,
                             weights = data$log_growth))
colnames(cumsum)<-c('Time','Group','Cumsum')
bar1<-ggbarplot(cumsum, x = "Time", y = "Cumsum", fill = "Group",
                label = TRUE,lab.pos = c("in"),
                lab.col = "red", lab.nb.digits=2,
                palette = "Paired", width=0.9)

#Cumulative sum of log-growth by Group according to Time
cumsum1<-data.frame(wtd.table(x = data$Group, y = data$time1,
                             weights = data$log_growth))
colnames(cumsum1)<-c('Group','Time','Cumsum')
bar2<-ggbarplot(cumsum1, x = "Group", y = "Cumsum", fill = "Time",label = TRUE,
                label = TRUE,lab.pos = c("in"),
                lab.col = "red", lab.nb.digits=2,
                palette = "Paired", width=0.9)
grid.arrange(bar1,bar2)
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Gráfico de barras apilado con suma acumulada

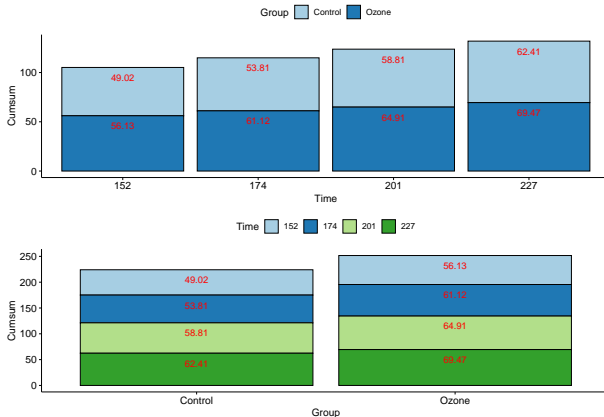


Figura 5: Perfiles individuales grupo Control. Abetos

**Gráfico de mosaico.** Un diagrama de mosaico (también conocido como diagrama de Marimekko<sup>5</sup> [The earliest known mosaic display goes back to Georg von Mayr (1877).]) es un método gráfico para visualizar datos de dos o más variables cualitativas. Es la extensión multidimensional de los diagramas de columna, que muestran gráficamente la misma información para una sola variable. Es un tipo especial de gráfico de barras apiladas que muestra porcentajes de datos en grupos. Es una representación gráfica de una tabla de contingencia.

---

<sup>5</sup>John Hartigan, Beat Kleiner: Mosaics for contingency tables. In: Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface. 1981, S. 268–273. Michael Friendly: A Brief History of the Mosaic Display In: Journal of Computational and Graphical Statistics, 2002, 11, 89–107.<https://www.datavis.ca/papers/moshist.pdf>

Los principales defectos de los diagramas de Marimekko son que pueden ser difíciles de leer, especialmente con un gran número de segmentos. Además, es difícil hacer comparaciones precisas entre cada segmento, ya que no están dispuestos uno junto al otro a lo largo de una línea de base común. Por lo tanto, los diagramas de Marimekko son mejores para dar una visión general de los datos.<sup>6</sup>

---

<sup>6</sup>https:

[//datavizcatalogue.com/ES/metodos/diagrama\\_de\\_marimekko.html](https://datavizcatalogue.com/ES/metodos/diagrama_de_marimekko.html)

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

## Gráfico de mosaico. Un diagrama de mosaico

```
library(ggplot2)
library(plyr)
library(reshape2)
library(questionr)
library(gridExtra)
library(vcd)
library(readxl)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$i..ID
data$Time<-as.factor(data$time1)
data2 <-wtd.table(x = data$Time, y=data$Group, weights = data$log_growth)
dimnames(data2) <- list(Time=attributes(data2)$dimnames[[1]],
                        Group=attributes(data2)$dimnames[[2]])

labs<-data2
mosaic(t(data2),
      split = TRUE,
      colorize = FALSE,
      shade = TRUE,
      pop = FALSE,
      gp = gpar(fill = matrix(c("blue", "blue", "purple","purple", "yellow", "yellow","red","red"), 2, 4),
labeling_cells(text = data2)(t(data2))
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de mosaico. Un diagrama de mosaico

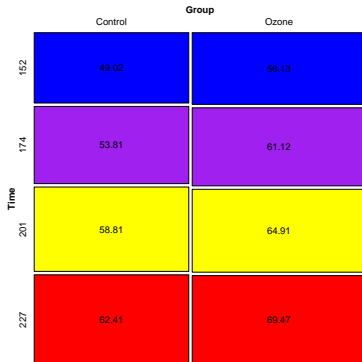


Figura 6: diagrama de Marimekko/Mosaico

**Spine plot o diagrama de columna.** Un diagrama de columna es un tipo específico de diagrama de mosaico para solo dos variables. Es similar al gráfico de barras compuestas, pero ahora el ancho de la barra se establece según la proporción de cada categoría horizontal.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Spine plot o diagrama de columna. Ejemplo de un spineplot usando los datos de abetos.

```
library(ggplot2)
library(plyr)
library(reshape2)
library(questionr)
library(gridExtra)
library(ggmosaic)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$í..ID
data$Time<-as.factor(data$time1)
data$Grupo<-as.factor(data$Group)

par(mfrow=c(1,2))

spineplot(Grupo~log_growth, data=data,col=c("blue","dark red"), breaks = quantile(data$log_growth),
          ylevels = 2:1)

spineplot(Time~log_growth, data=data,col=c("blue","dark red","dark magenta", "yellow"),
          breaks = quantile(data$log_growth),
          ylevels = 1:4)
```



# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Spine plot o diagrama de columna. Ejemplo de un spineplot usando los datos de abetos.

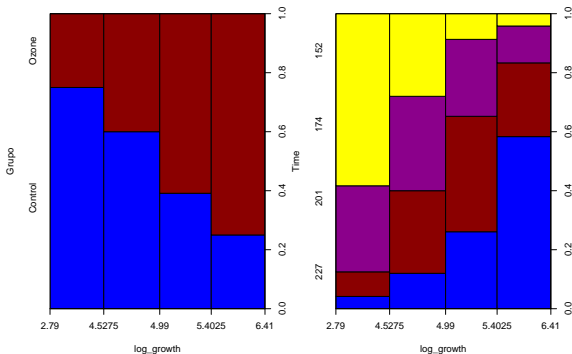


Figura 7: Spineplot

**Density plot o gráfico de densidades**<sup>7</sup>. Un gráfico de densidad es una representación de la distribución de una variable numérica. Utiliza una estimación de densidad kernel para mostrar la función de densidad de probabilidad de la variable (ver más). Es una versión suavizada del histograma y se usa en el mismo concepto.

---

<sup>7</sup>Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*. 33 (3): 1065–1076. doi:10.1214/aoms/1177704472. JSTOR 2237880.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Density plot o gráfico de densidades.** Por ejemplo, un Density plot o gráfico de densidades global (ignorando la naturaleza longitudinal de los datos) de los logaritmos de crecimiento de acuerdo al grupo se puede obtener con:

```
#Global density plot ignoring longitudinality
library(ggplot2)
data<-read.csv(file="SPRUCE1.csv",header=T,
               sep=',',dec='.')
data$ID<-data$ID
means<-round(c(by(data$log_growth,data$Group,mean)),2)
ggplot(data, aes(x=log_growth,col=Group,fill=Group,group=Group)) +
  geom_density(alpha=0.5)+
  geom_vline(xintercept=means[1], size=1.2, color="tomato")+
  geom_vline(xintercept=means[2], size=1.2, color="cyan")+
  labs(fill = "Group",y="Probability",x="Log-Growth")+
  geom_text(aes(x=means[1]-0.2,label=paste0("Mean\n",means[1]), y=0.6),size=5,col="tomato")+
  geom_text(aes(x=means[2]+0.3,label=paste0("Mean\n",means[2]), y=0.6),size=5,col="blue")
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Density plot o gráfico de densidades.** Por ejemplo, un Density plot o gráfico de densidades global (ignorando la naturaleza longitudinal de los datos) de los logaritmos de crecimiento de acuerdo al grupo se puede obtener con:

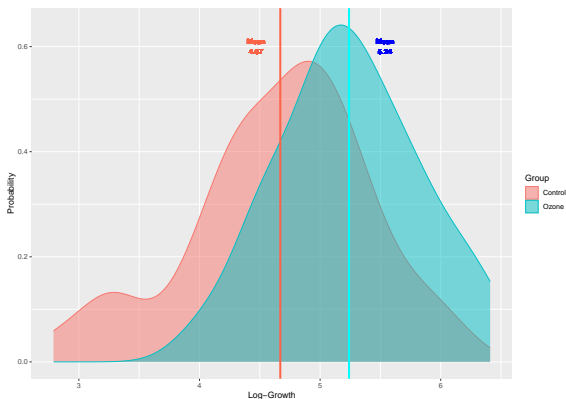


Figura 8: Density plot. Abetos

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Density plot o gráfico de densidades.** Por ejemplo, un Density plot o gráfico de densidades global (ignorando la naturaleza longitudinal de los datos) de los logaritmos de crecimiento de acuerdo al tiempo se puede obtener con:

```
#Global density plot ignoring longitudinality
library(ggplot2)
data<-read.csv(file="SPRUCE1.csv",header=T,
               sep=',',dec='.')
data$ID<-data$ï..ID
data$Time<-as.factor(data$time1)

means<-round(c(by(data$log_growth,data$Time,mean)),2)
ggplot(data, aes(x=log_growth,col=Time,fill=Time,group=Time)) +
  geom_density(alpha=0.5)+
  geom_vline(xintercept=means[1], size=1, color="tomato")+
  geom_vline(xintercept=means[2], size=1, color="cyan")+
  geom_vline(xintercept=means[3], size=1, color="blue")+
  geom_vline(xintercept=means[4], size=1, color="purple")+
  labs(fill = "Time",y="Probability",x="Log-Growth")+
  geom_text(aes(x=means[1],label=paste0("Mean\n",means[1]), y=0.6),size=5,col="tomato")+
  geom_text(aes(x=means[2],label=paste0("Mean\n",means[2]), y=0.6),size=5,col="cyan")+
  geom_text(aes(x=means[3],label=paste0("Mean\n",means[3]), y=0.6),size=5,col="blue")+
  geom_text(aes(x=means[4],label=paste0("Mean\n",means[4]), y=0.6),size=5,col="purple")
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

**Density plot o gráfico de densidades.** Por ejemplo, un Density plot o gráfico de densidades global (ignorando la naturaleza longitudinal de los datos) de los logaritmos de crecimiento de acuerdo al tiempo se puede obtener con:

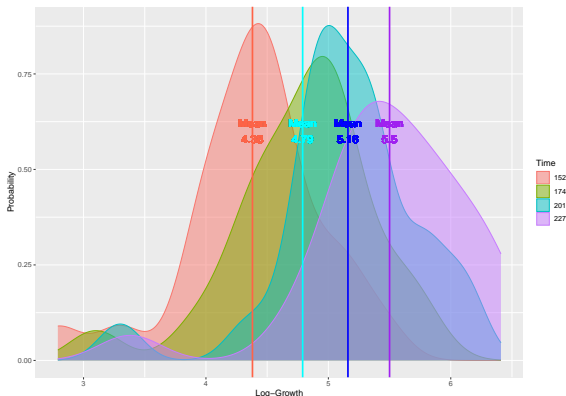


Figura 9: Density plot. Abetos

**Ridgelineplot o gráfico de crestas.** Un Ridgelineplot (anteriormente llamado Joyplot) permite estudiar la distribución de una variable numérica para varios grupos. Un gráfico de crestas es simplemente un conjunto de muchos gráficos de densidad.

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ridgelineplot o gráfico de crestas. Ejemplo de un ridgelineplot de acuerdo a la variable categórica “Group”. Abetos:

```
library(ggribes)
library(ggplot2)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$i..ID
data$Time<-as.factor(data$time1)

ggplot(data, aes(x = log_growth, y = Group, fill = Group)) +
  geom_density_ridges(quantile_lines=TRUE,
    quantile_fun=function(log_growth,...)mean(log_growth),alpha=0.7) +
  theme_ridges() +
  theme(legend.position = "none")
```



# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ridgelineplot o gráfico de crestas. Ejemplo de un ridgelineplot de acuerdo a la variable categórica "Group". Abetos:

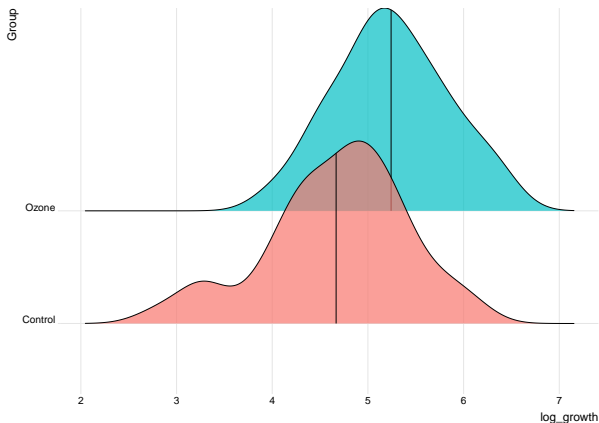


Figura 10: Ridgeline plot de ambos grupos. Abetos

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ridgelineplot o gráfico de crestas. Ejemplo de un ridgelineplot de acuerdo a la variable categórica “Time”. Abetos:

```
library(ggribes)
library(ggplot2)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$i..ID
data$Time<-as.factor(data$time1)

ggplot(data, aes(x = log_growth, y = Time, fill = Time)) +
  geom_density_ridges(quantile_lines=TRUE,
    quantile_fun=function(log_growth,...)mean(log_growth),alpha=0.7) +
  theme_ridges() +
  theme(legend.position = "none")
```

# ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ridgelineplot o gráfico de crestas. Ejemplo de un ridgelineplot de acuerdo a la variable categórica "Time". Abetos:

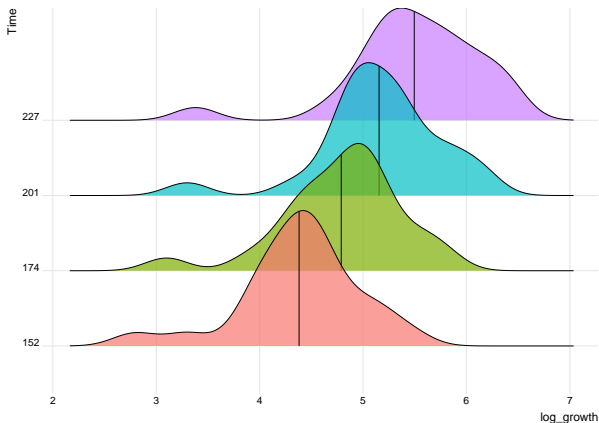


Figura 11: Ridgeline plot de acuerdo al tiempo. Abetos