

Análisis Físico Químico

2024-06-24

Análisis físico-químico de las muestras

```
variablesFisicas <- read.csv('fisicoQuimico.csv', encoding = 'latin1', sep = ';')
```

Realiza el test de normalidad de shapiro para cada una de las variables

```
variablesFisicas %>%  
  summarise(p_valor_temperatura = shapiro.test(temp.C.)$p.value,  
            p_valor_oxigenoDisuelto = shapiro.test(Odisuelto..mg.L.)$p.value,  
            p_valor_conductividadElectrica = shapiro.test(condElec.µS.cm.)$p.value,  
            p_valor_pH = shapiro.test(pH)$p.value,  
            p_valor_dureza = shapiro.test(dureza.mg.LCaCO3.)$p.value,  
            p_valor_caudal = shapiro.test(caudal.m3.s.1.)$p.value)
```

```
##  p_valor_temperatura p_valor_oxigenoDisuelto p_valor_conductividadElectrica  
## 1          0.6560092          0.2639086          0.4718902  
##  p_valor_pH p_valor_dureza p_valor_caudal  
## 1  0.8730672    0.001797036    3.418642e-05
```

Según los resultados, solo el caudal y la dureza se dsitribuyen normalmente, ya que tienen p valores menores que 0.05 ($p < 0.05$). Para comprobar, se realiza el test de Anderson-Darling

```
variablesFisicas %>%  
  summarise(p_valor_temperatura = ad.test(temp.C.)$p.value,  
            p_valor_oxigenoDisuelto = ad.test(Odisuelto..mg.L.)$p.value,  
            p_valor_conductividadElectrica = ad.test(condElec.µS.cm.)$p.value,  
            p_valor_pH = ad.test(pH)$p.value,  
            p_valor_dureza = ad.test(dureza.mg.LCaCO3.)$p.value,  
            p_valor_caudal = ad.test(caudal.m3.s.1.)$p.value)
```

```
##  p_valor_temperatura p_valor_oxigenoDisuelto p_valor_conductividadElectrica  
## 1          0.5615071          0.1641355          0.497706  
##  p_valor_pH p_valor_dureza p_valor_caudal  
## 1  0.8922803    0.0041493    1.162537e-05
```

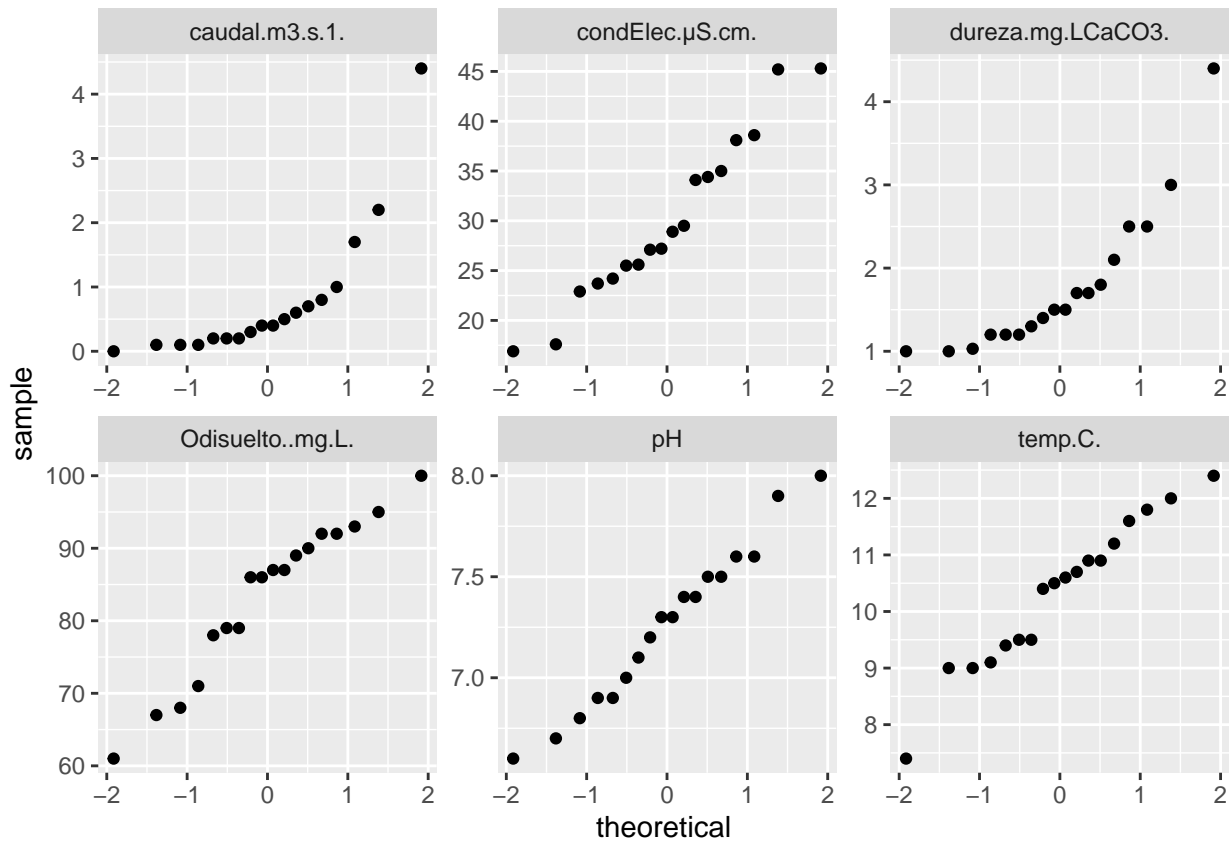
Al realizar el test de Anderson-Darling, se comprueba que el caudal y la dureza son las unicas variables que se distribuyen normamlente.

Mostramos gráficamente la Distribución de los datos

Grafica el Q-Q plot con ggplot2

```
variablesFisicas %>%  
  select(c(4:9)) %>%  
  gather(variable, valor) %>%
```

```
ggplot(aes(sample = valor)) +
  stat_qq() +
  facet_wrap(~ variable, scales = 'free')
```



Graficamente se confirma que el caudal y la dureza cumplen con distribuirse normalmente.

Estandarización de las variable a $\log(x + 1)$ ya que hay una observación con 0

Estandarización logarítmica para cada variable

```
variablesEstandarizadas <- variablesFisicas %>%
  mutate(
    temp.C. = log(temp.C. + 1),
    Odisuelto..mg.L. = log(Odisuelto..mg.L. + 1),
    condElec.µS.cm. = log(condElec.µS.cm. + 1),
    pH = log(pH + 1),
    dureza.mg.LCaCO3. = log(dureza.mg.LCaCO3. + 1),
    caudal.m3.s.1. = log(caudal.m3.s.1. + 1)
  )
```

Análisis de componentes principales por río

Escalado de las variables

```
escalado <- scale(variablesEstandarizadas[,4:9])
escalado <- cbind(escalado, variablesEstandarizadas[,1:3])
```

Análisis de componentes para la cuenca Penagos

```
penagos <- escalado %>%
  filter(cuenca == 'Penagos')
```

Aplica el pca

```
penagos.pca <- prcomp(penagos[, 1:6], scale = TRUE)
```

Resultados del análisis

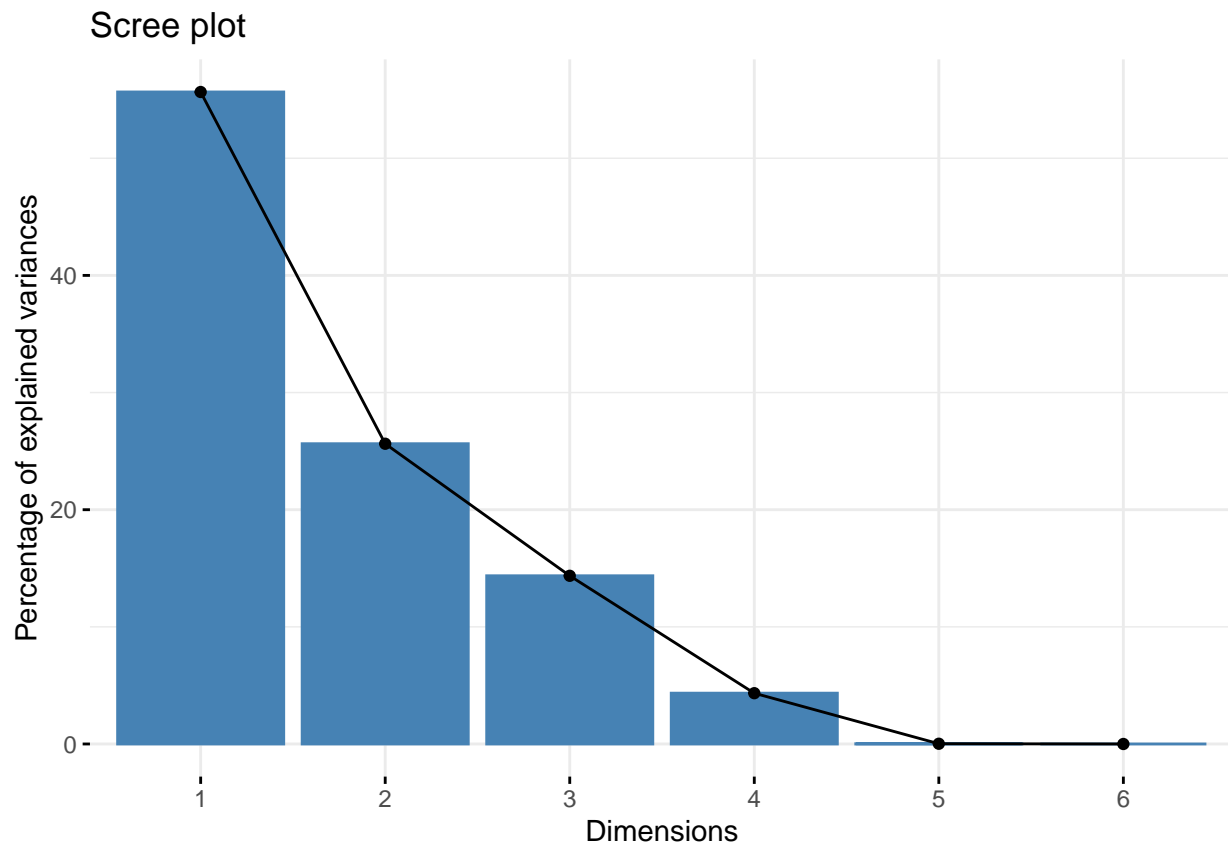
```
summary(penagos.pca)
```

```
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6
## Standard deviation    1.8274 1.2401 0.9281 0.51008 0.03555 2.697e-17
## Proportion of Variance 0.5565 0.2563 0.1436 0.04336 0.00021 0.000e+00
## Cumulative Proportion 0.5565 0.8128 0.9564 0.99979 1.00000 1.000e+00
```

Se observa que la primera componente explica el 55.65% de la varianza total, mientras que las primeras tres explican el 95.64%

Visualización gráfica

```
fviz_eig(penagos.pca)
```



Distribución de las observaciones. Observaciones similares estan mas cercanas

```
fviz_pca_ind(penagos.pca,  
  col.ind = "cos2", # Color by the quality of representation  
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
  repel = TRUE      # Avoid text overlapping  
)
```

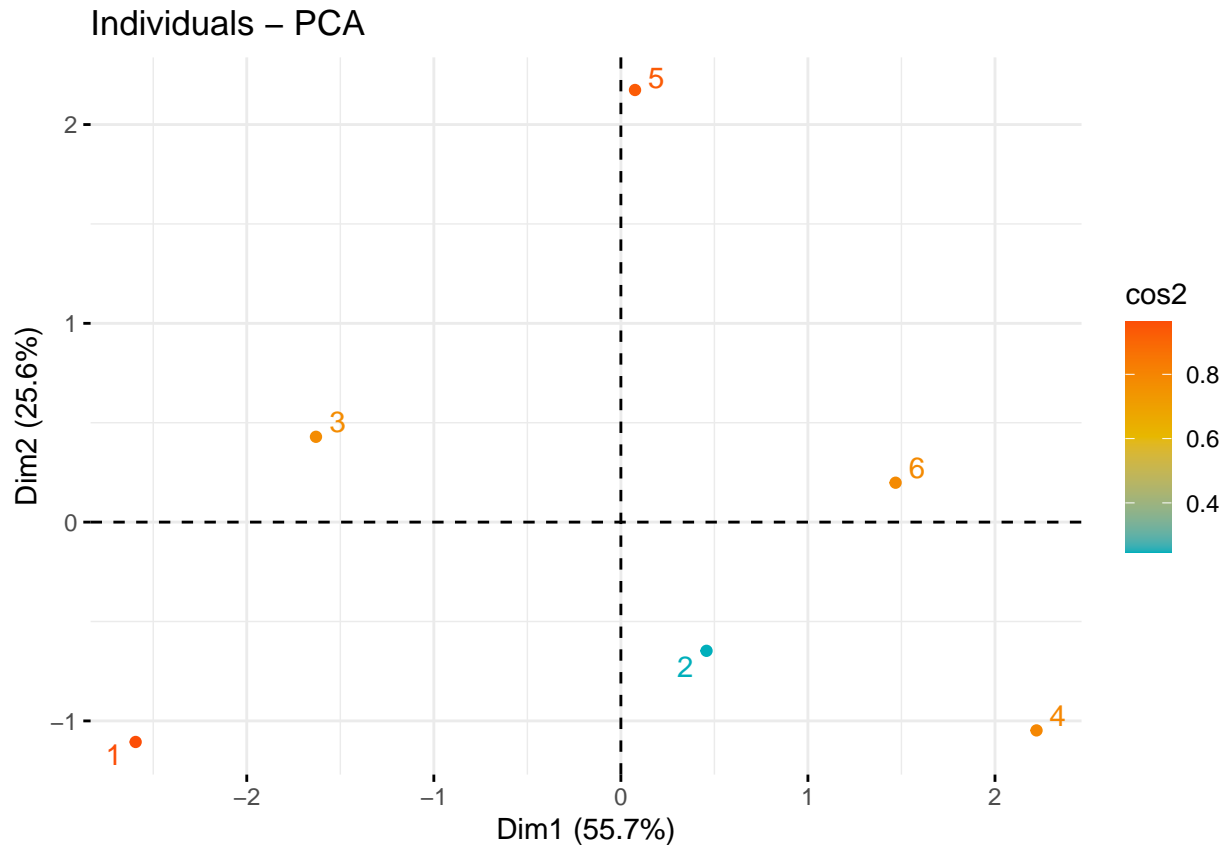
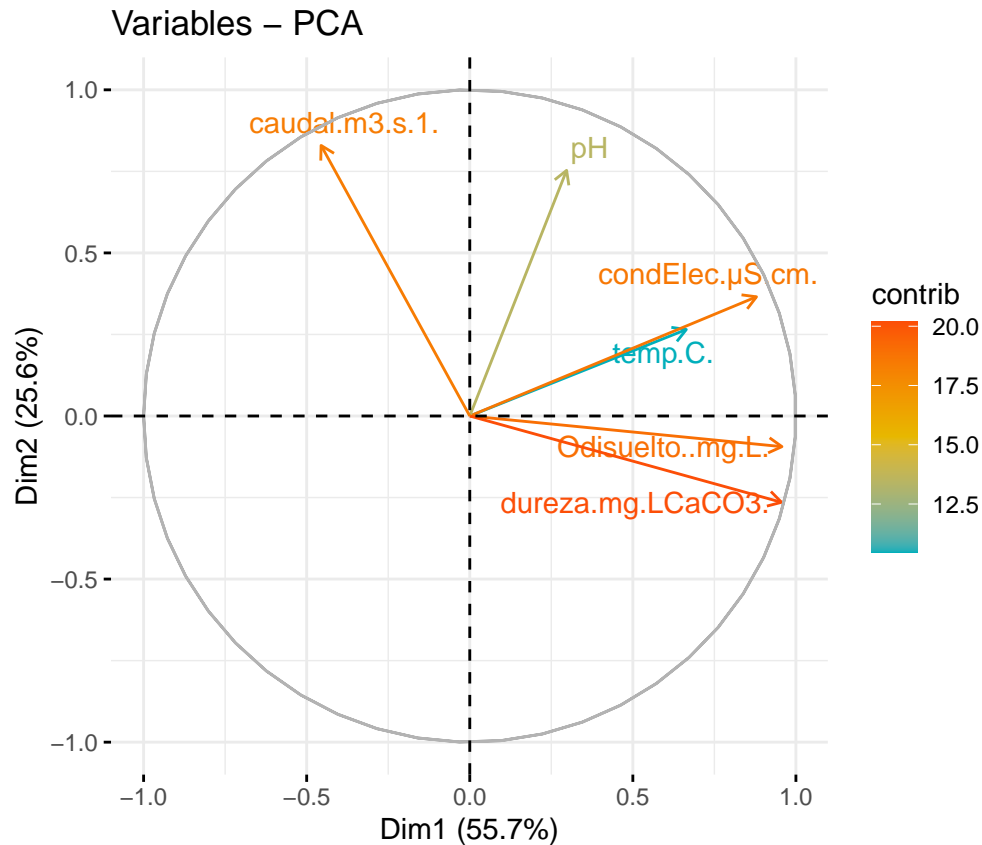


Gráfico de las variables. Las que correlacionan positivamente se encuentran en el mismo cuadrante de la gráfica. Las que correlacionan negativamente se encuentran en cuadrantes opuestos. Por ejemplo, caudal correlacionan negativamente con la dureza y el oxígeno disuelto.

El color representa la contribución a la componente. Las variables de la temperatura y el pH son las que menos contribuyen a las componentes.

```
fviz_pca_var(penagos.pca,
  col.var = "contrib", # Color by contributions to the PC
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE        # Avoid text overlapping
)
```



Muestra cómo se agrupan las observaciones por periodo

```
groups <- as.factor(penagos$Periodo)
periodo <- fviz_pca_ind(penagos.pca,
  col.ind = groups, # color by groups
  palette = c("#00AFBB", "#FC4E07"),
  addEllipses = TRUE, # Concentration ellipses
  ellipse.type = "confidence",
  legend.title = "C poca",
  repel = TRUE
)
```

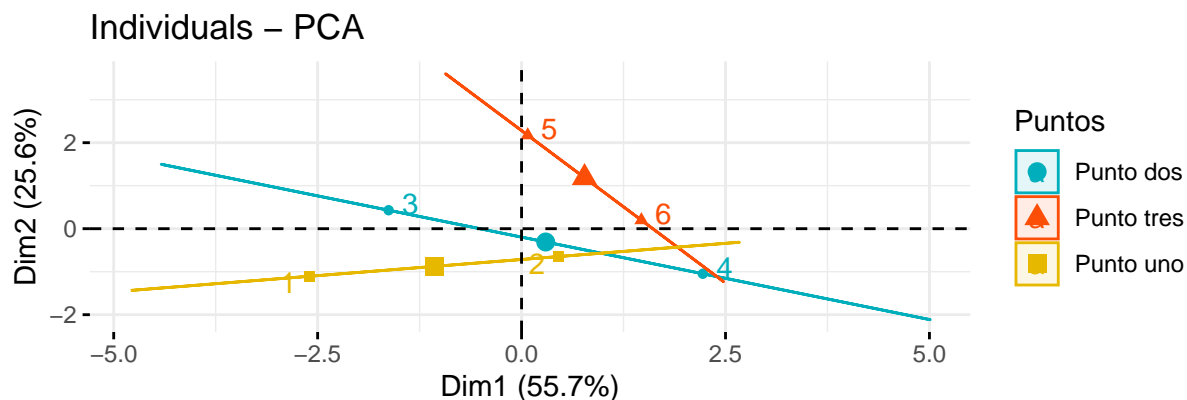
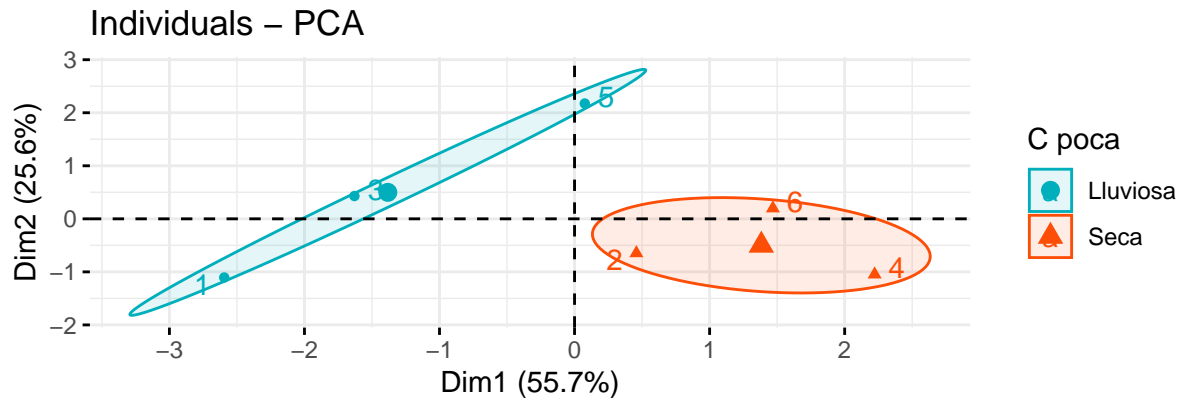
Muestra cómo se agrupan las observaciones por punto

```
groups <- as.factor(penagos$punto)
punto <- fviz_pca_ind(penagos.pca,
  col.ind = groups, # color by groups
  palette = c("#00AFBB", "#FC4E07", "#E7B800"),
  addEllipses = TRUE, # Concentration ellipses
  ellipse.type = "confidence",
  legend.title = "Puntos",
  repel = TRUE
)
```

Muestra las dos gráficas juntas

```
periodo/punto
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x09 in encoding cp1252
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x09 in encoding cp1252
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x09 in encoding cp1252
```



Análisis de componentes para la cuenca Morro

```
morro <- escalado %>%
  filter(cuenca == 'Morro')
```

Aplica el pca

```
morro.pca <- prcomp(morro[, 1:6], scale = TRUE)
```

Resultados del análisis

```
summary(morro.pca)
```

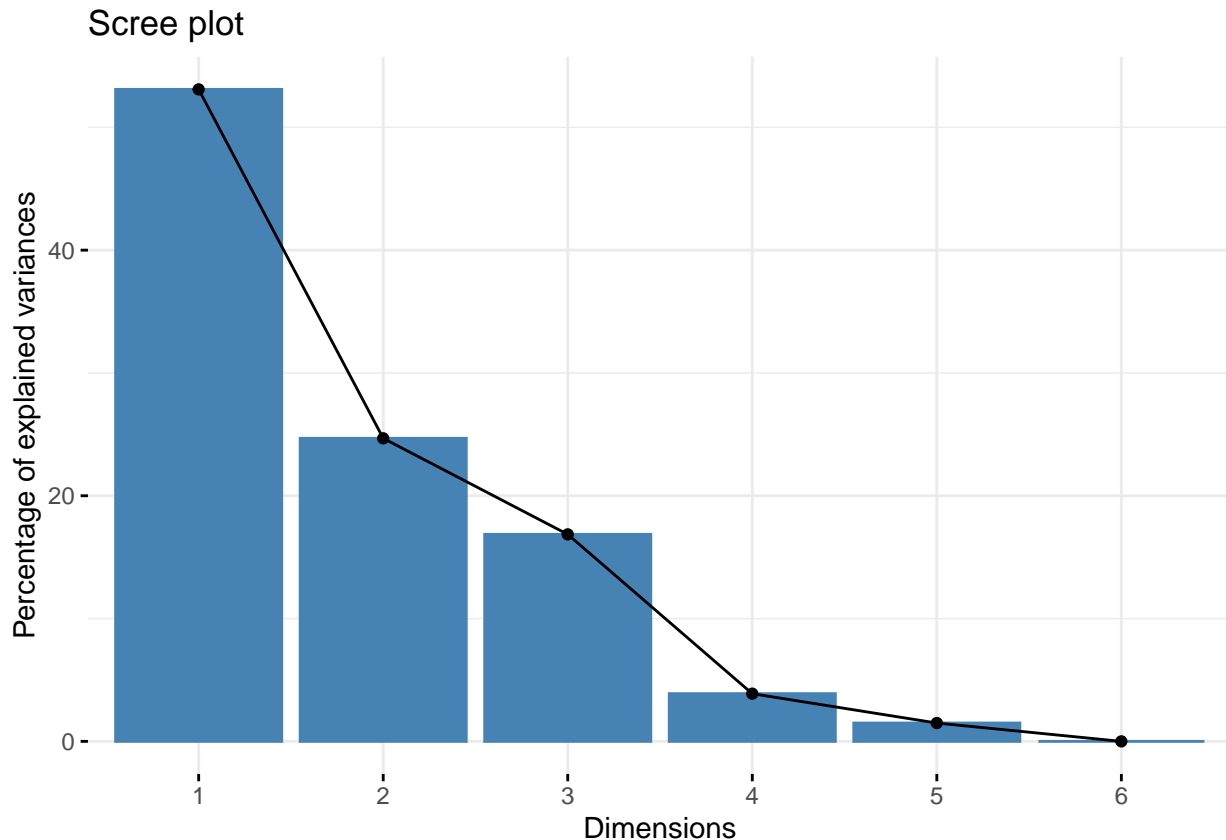
```
## Importance of components:
##          PC1      PC2      PC3      PC4      PC5      PC6
```

```
## Standard deviation      1.7848 1.2168 1.0057 0.48276 0.29931 4.25e-17
## Proportion of Variance 0.5309 0.2468 0.1686 0.03884 0.01493 0.00e+00
## Cumulative Proportion  0.5309 0.7777 0.9462 0.98507 1.00000 1.00e+00
```

Se observa que la primera componente explica el 53.09% de la varianza total, mientras que las primeras tres explican el 94.62%. Obsérvese que dos componentes explican mas de 3 cuartas partes de la varianza total

Visualización gráfica

```
fviz_eig(morro.pca)
```



Distribución de las observaciones. Observaciones similares estan mas cercanas

```
fviz_pca_ind(morro.pca,
  col.ind = "cos2", # Color by the quality of representation
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE      # Avoid text overlapping
)
```

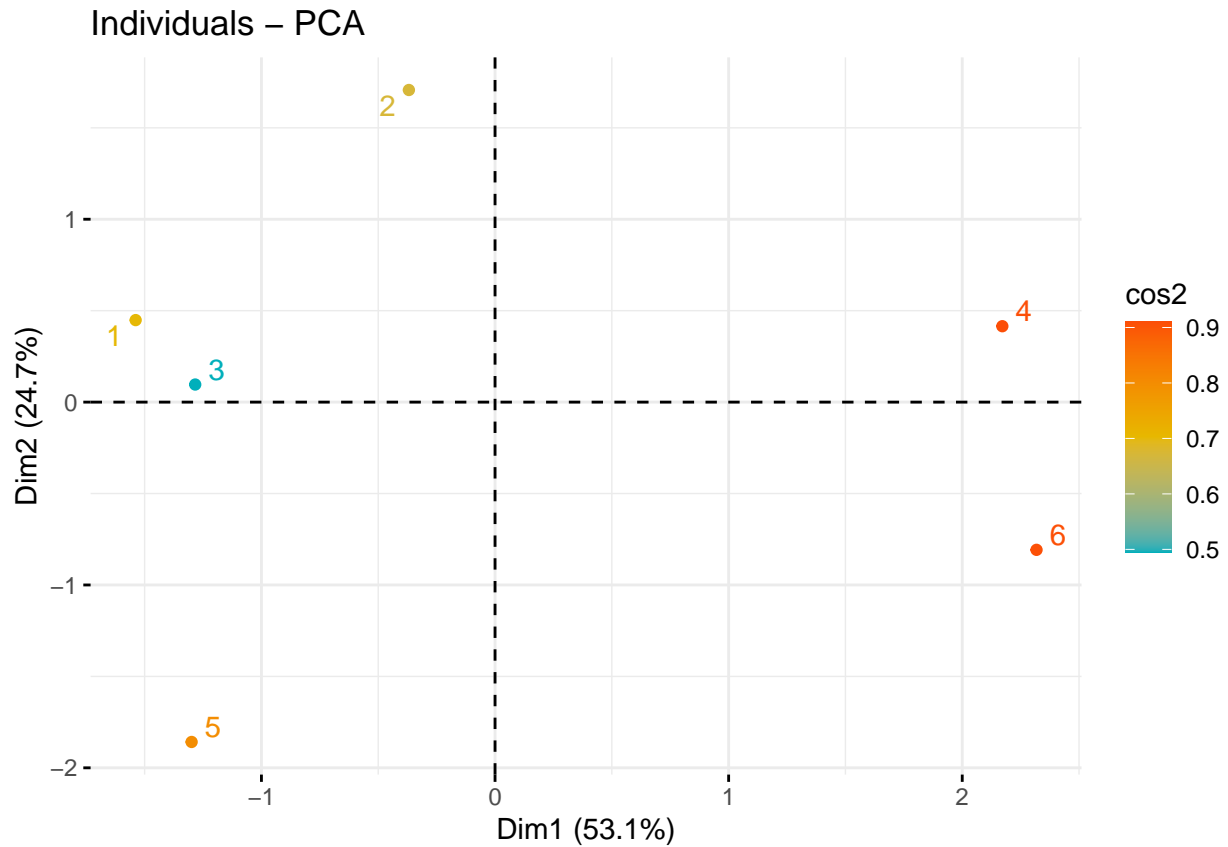
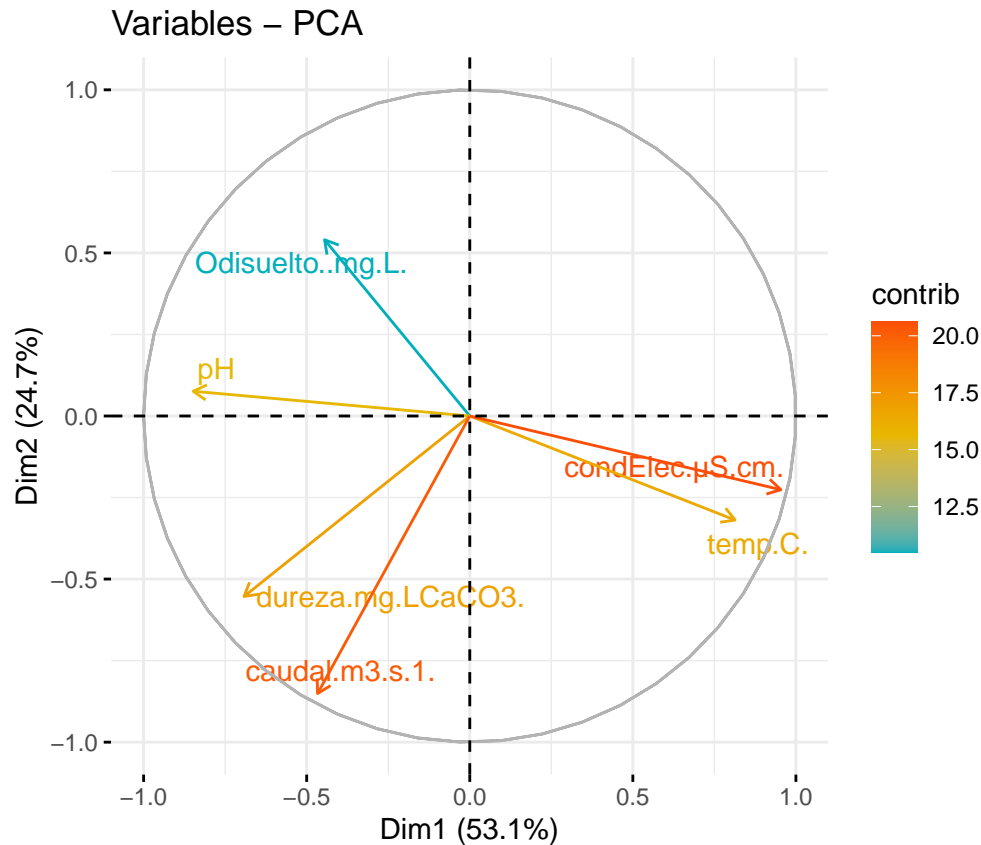



Gráfico de las variables. Las que correlacionan positivamente se encuentran en el mismo cuadrante de la gráfica. Las que correlacionan negativamente se encuentran en cuadrantes opuestos.

El color representa la contribución a la componente. La variable del oxígeno disuelto es la que menos contribuye a las componentes.

```
fviz_pca_var(morro.pca,
  col.var = "contrib", # Color by contributions to the PC
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE        # Avoid text overlapping
)
```



Muestra cómo se agrupan las observaciones por periodo

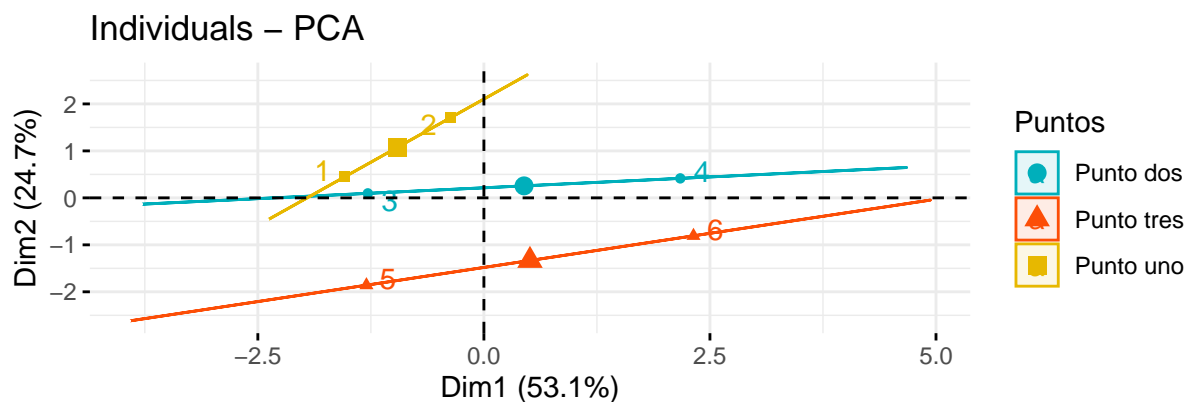
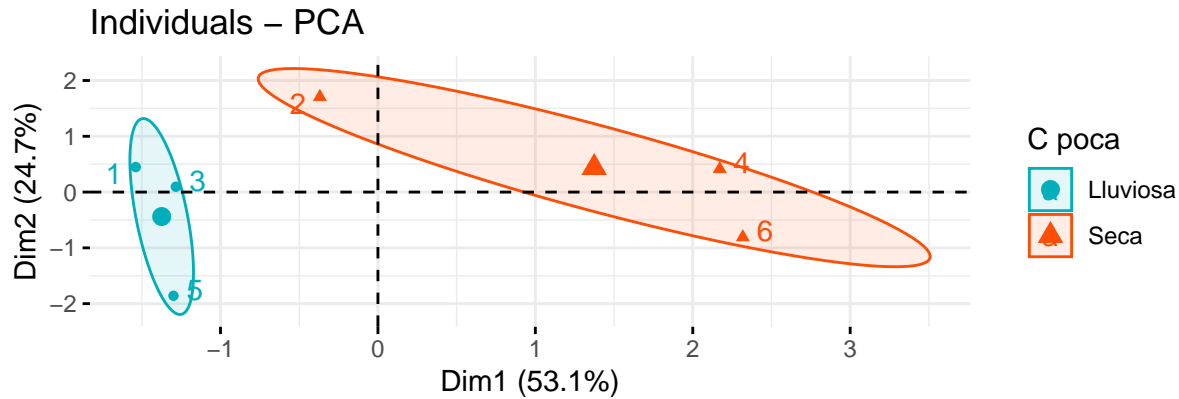
```
groups <- as.factor(morro$Periodo)
periodo <- fviz_pca_ind(morro.pca,
  col.ind = groups, # color by groups
  palette = c("#00AFBB", "#FC4E07"),
  addEllipses = TRUE, # Concentration ellipses
  ellipse.type = "confidence",
  legend.title = "C poca",
  repel = TRUE
)
```

Muestra cómo se agrupan las observaciones por punto

```
groups <- as.factor(morro$punto)
punto <- fviz_pca_ind(morro.pca,
  col.ind = groups, # color by groups
  palette = c("#00AFBB", "#FC4E07", "#E7B800"),
  addEllipses = TRUE, # Concentration ellipses
  ellipse.type = "confidence",
  legend.title = "Puntos",
  repel = TRUE
)
periodo/punto
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
```

```
## width unknown for character 0x09 in encoding cp1252
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x09 in encoding cp1252
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x09 in encoding cp1252
```



Análisis de componentes para la cuenca Tejar

```
tejar <- escalado %>%
  filter(cuenca == 'Tejar ')
```

Aplica el pca

```
tejar.pca <- prcomp(tejar[, 1:6], scale = TRUE)
```

Resultados del análisis

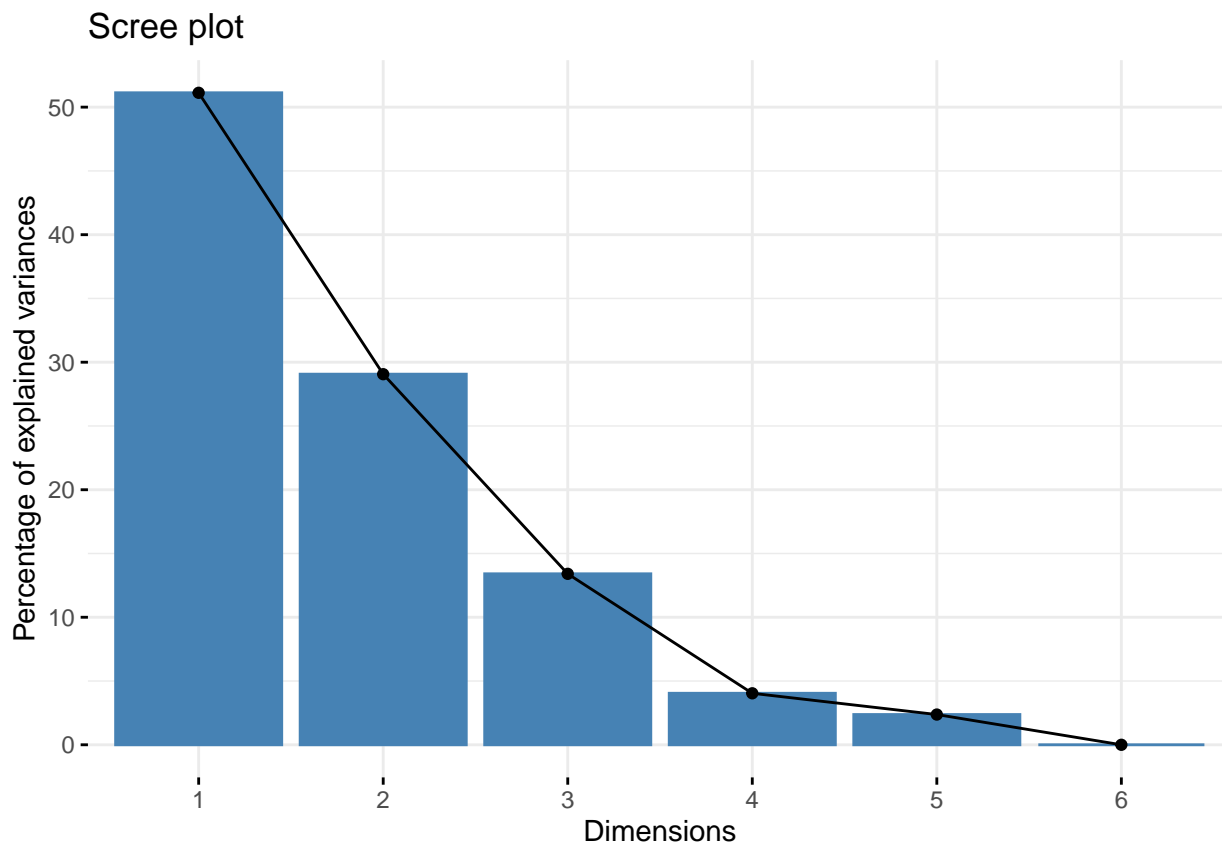
```
summary(tejar.pca)
```

```
## Importance of components:
##              PC1    PC2    PC3    PC4    PC5    PC6
## Standard deviation 1.7514 1.3205 0.8968 0.49222 0.37727 9.018e-17
## Proportion of Variance 0.5112 0.2906 0.1341 0.04038 0.02372 0.000e+00
## Cumulative Proportion 0.5112 0.8018 0.9359 0.97628 1.00000 1.000e+00
```

Se observa que la primera componente explica el 51.12% de la varianza total, mientras que las primeras tres explican el 93.59%

Visualización gráfica

```
fviz_eig(tejar.pca)
```



Distribución de las observaciones. Observaciones similares están más cercanas

```
fviz_pca_ind(tejar.pca,  
  col.ind = "cos2", # Color by the quality of representation  
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
  repel = TRUE      # Avoid text overlapping  
)
```

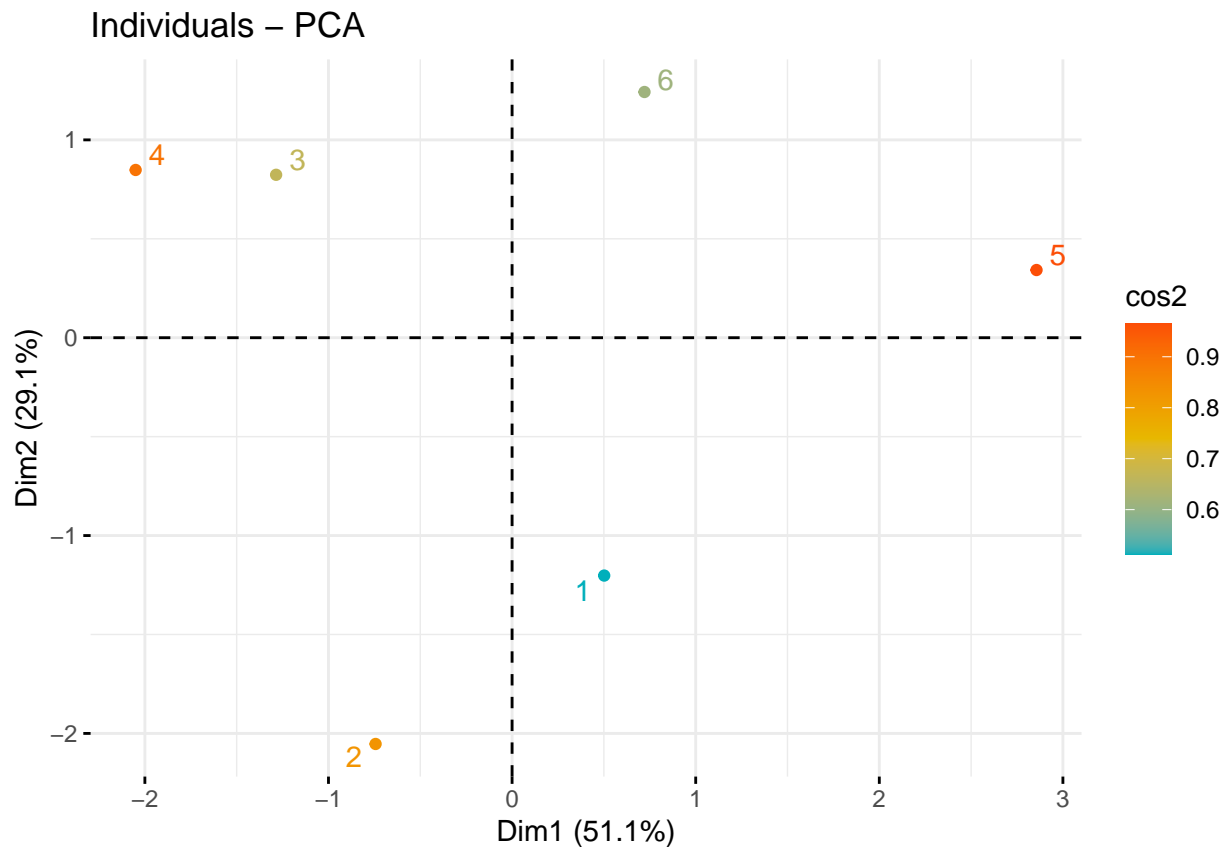
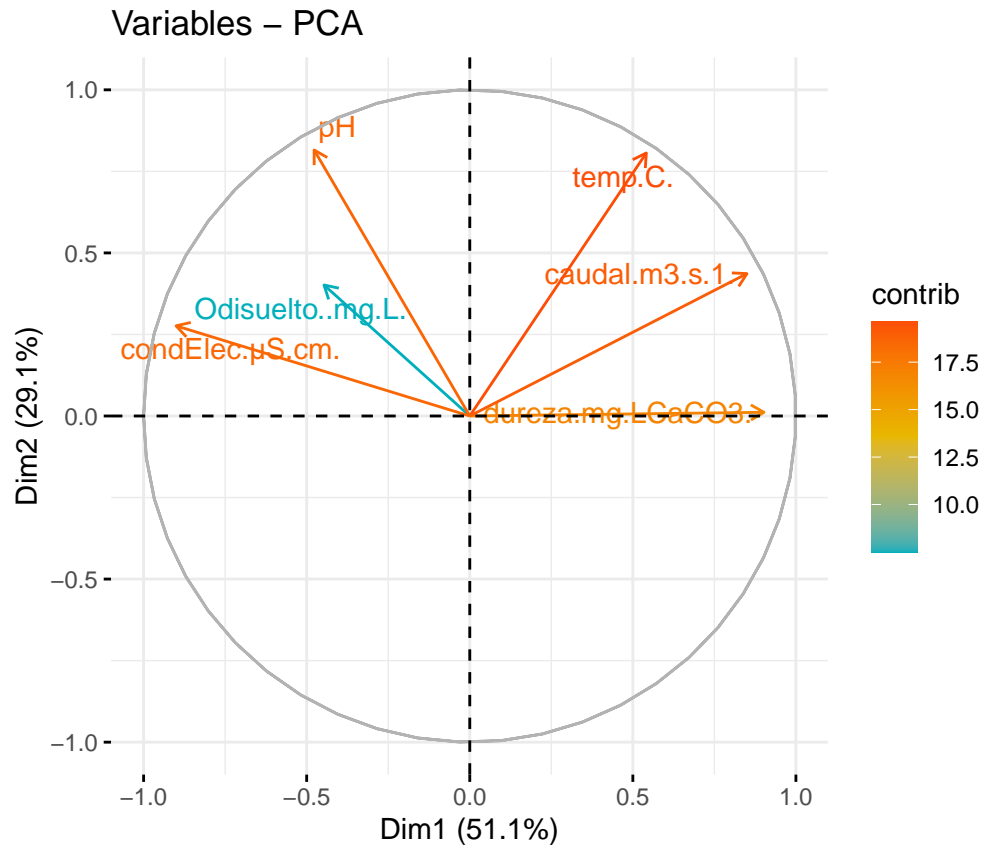


Gráfico de las variables. Las que correlacionan positivamente se encuentran en el mismo cuadrante de la gráfica. Las que correlacionan negativamente se encuentran en cuadrantes opuestos. Por ejemplo, caudal correlacionan negativamente con la dureza y el oxígeno disuelto.

El color representa la contribución a la componente. La variable del oxígeno disuelto es la que menos contribuye a las componentes.

```
fviz_pca_var(tejar.pca,
  col.var = "contrib", # Color by contributions to the PC
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE        # Avoid text overlapping
)
```



Muestra cómo se agrupan las observaciones por periodo

```
groups <- as.factor(tejar$Periodo)
periodo <- fviz_pca_ind(tejar.pca,
  col.ind = groups, # color by groups
  palette = c("#00AFBB", "#FC4E07"),
  addEllipses = TRUE, # Concentration ellipses
  ellipse.type = "confidence",
  legend.title = "C poca",
  repel = TRUE
)
```

Muestra cómo se agrupan las observaciones por punto

```
groups <- as.factor(penagos$punto)
punto <- fviz_pca_ind(tejar.pca,
  col.ind = groups, # color by groups
  palette = c("#00AFBB", "#FC4E07", "#E7B800"),
  addEllipses = TRUE, # Concentration ellipses
  ellipse.type = "confidence",
  legend.title = "Puntos",
  repel = TRUE
)
periodo/punto
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
```

```
## width unknown for character 0x09 in encoding cp1252
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x09 in encoding cp1252
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x09 in encoding cp1252
```

