



## Análisis de componentes principales aplicado a un indicador de corporaciones tecnológicas

Pérez, R.S.<sup>1</sup>      Bermann, M.A.<sup>2</sup>

17 diciembre, 2021

<sup>1</sup>Rafael Sergio Pérez Pujol, UCLM, [RafaelSergio.Perez@alu.uclm.es](mailto:RafaelSergio.Perez@alu.uclm.es)

<sup>2</sup>Mateo Alberto Bermann Albalat, UCLM, [MateoAlberto.Bermann@alu.uclm.es](mailto:MateoAlberto.Bermann@alu.uclm.es)

# Índice general

<b>Resumen</b>	<b>2</b>
<b>1. Introducción</b>	<b>3</b>
<b>2. Análisis de correlaciones</b>	<b>5</b>
<b>3. Análisis descriptivo multivariante</b>	<b>7</b>
<b>4. Análisis de componentes</b>	<b>9</b>
4.1. Ejecución del análisis ACP . . . . .	9
4.2. Componentes a retener . . . . .	11
4.3. Interpretación de las componentes principales analizadas . . . . .	12
<b>5. Presentación de resultados y conclusiones</b>	<b>13</b>
<b>Referencias bibliográficas</b>	<b>18</b>
<b>Anexos</b>	<b>20</b>
Anexo 1. Datos de la sesión . . . . .	20
Anexo 2. Base de datos . . . . .	22
Anexo 3. Código ( <i>script</i> ) utilizado . . . . .	23

# Resumen

Las grandes corporaciones tecnológicas han sufrido un crecimiento exponencial en el siglo XXI que se ha visto impulsado en mayor medida por el auge cada vez más notable de los medios digitales y las nuevas plataformas y tendencias comunicativas y de ocio digital que protagonizan el actual escenario. Así, empresas como Apple, Google (Alphabet) o Tesla han mejorado de forma considerable su posición en los mercados financieros en los últimos 15 años. Con ello, surgen indicadores que intentan recoger, de forma sectorial, los valores bursátiles para mejorar los mecanismos de reasignación de recursos financieros. Es por ello por lo que, en el marco del Máster Universitario en Modelización y Análisis de Datos Económicos, y en concreto en el área de Técnicas Multivariantes Aplicadas al Análisis Sectorial, se va a plantear el desarrollo de un informe,<sup>1</sup> basado en un análisis de componentes principales para emular el índice NASDAQ 100 ( $\hat{NDX}$ ) a partir de diferentes valores cotizados del propio índice.

---

<sup>1</sup>Este informe ha sido realizado con el software R, a través del entorno RStudio y se ha maquetado, mediante R Markdown, a partir de las ayudas de libros, artículos y clases de Allaire et al. (2021), AprendeR: Parte II (2021), Casero y Quintana (2021), Cano (2021), CRAN R-Project (2021), DataCamp (2021), Fernández (2021), Hlavac (2018), Keyes (2019), Kobi (2010), Luque (2019b), Luque (2019a), Tarancón y Fernández (2021), Van Hespén (2016), Xie, Dervieux, y Riederer (2021), Xie, Allaire, y Grolemond (2021) y Zhu (2020). También se ha tenido que recurrir a la instalación de MiKTeX (<https://miktex.org/>). Los datos se han importado de un archivo Microsoft® Excel®.

# 1. Introducción

A menudo, es fácil perderse en un mar de información financiera, y a veces acabamos con la sensación de que hay datos que complican el análisis en vez de dar información. En estos casos, nos interesa más la tendencia general que ir viendo datos empresa por empresa. Normalmente un índice, en sí mismo nos puede dar una ligera idea, pero sufre de varios problemas al ser un agregado, como el impacto de los *outliers* o el hecho de que integre empresas que vayan a una dirección completamente distinta a la tendencia del mercado.

Es por ello por lo que, junto a esto, y en el marco del análisis ACP, el **objetivo principal** de este informe pasa por determinar si una componente tendría la capacidad de explicar la mayor parte del comportamiento de los valores cotizados seleccionados, todo ello a través de un análisis de componentes principales donde, siguiendo a Fernández (2021), dichas componentes se basan en *“combinaciones lineales de las originales y se derivan en orden de importancia, de tal manera que la primera componente principal recoge, de la variación total de los datos originales, la mayor parte posible, y así sucesivamente”*, y donde el objetivo es *“ver si unas pocas componentes recogen la mayor parte de la variación de los datos originales. Si es así, se puede argüir que la dimensionalidad del problema no es  $p$  sino inferior a  $p$ ”*.

Para conseguir el objetivo final se ha recurrido a la base de datos de Yahoo (2021), a través del portal Yahoo! Finance,<sup>2</sup> extrayendo datos para 10 valores cotizados del NASDAQ 100 y el propio índice NASDAQ 100. El período de observación incluye todo el año 2019 y se han obtenido los valores de los rendimientos diarios que se han trimestralizado.

En este punto, cabe decir que el análisis de componentes principales requiere que el conjunto de datos con el que se trabaja esté correlacionado, pues el objetivo final es reducir la dimensionalidad de los datos y simplificarlos. Para ello, se recurrirá a un análisis de correlaciones en el siguiente capítulo.

Podemos ver de forma preliminar los datos de las 6 primeras observaciones (datos completos en el Anexo 2):

---

<sup>2</sup>La fuente de la base de datos es Yahoo (2021).

Cuadro 1: Vista inicial de los datos de las 6 primeras observaciones

	<b>T1</b>	<b>T2</b>	<b>T3</b>	<b>T4</b>
NDX	0.1	0.0	0.0	0.1
AAPL	0.2	0.0	0.1	0.3
GOOG	0.1	-0.1	0.1	0.1
MRVL	0.2	0.2	0.0	0.1
TSLA	-0.1	-0.3	0.1	0.5
SBUX	0.2	0.1	0.0	0.0

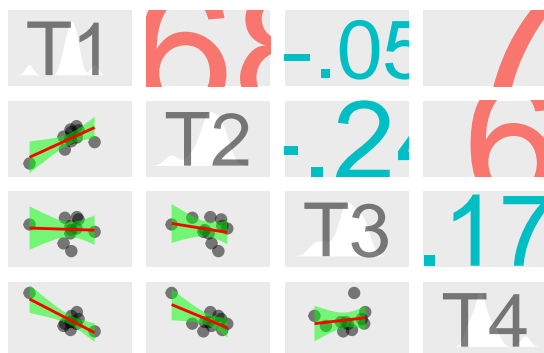
## 2. Análisis de correlaciones

Tal y como se ha comentado anteriormente, el análisis de componentes principales requiere unos valores de correlaciones elevados para el conjunto de datos analizados. Así, aplicando dicha premisa para el conjunto de datos utilizados para este informe, se puede obtener la matriz de correlaciones correspondiente:

Cuadro 2: Matriz de correlaciones valores e índice NASDAQ 100

	T1	T2	T3	T4
T1	1.00	0.68	-0.05	-0.77
T2	0.68	1.00	-0.24	-0.62
T3	-0.05	-0.24	1.00	0.17
T4	-0.77	-0.62	0.17	1.00

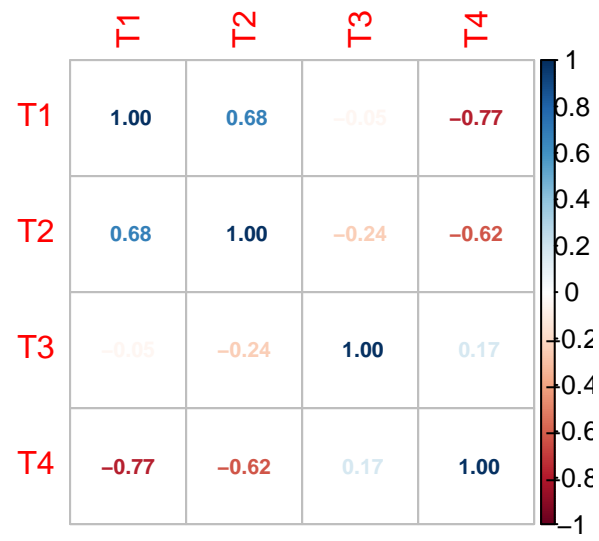
Así mismo, dicho análisis puede plantearse de forma gráfica a través de distintas opciones. En primer lugar, si se grafican las correlaciones con la librería `ez()`, se puede observar que dichos valores son relativamente elevados, especialmente para los conjuntos de T1-T2, T1-T4, y T2-T4.



De la misma forma, podemos observar la misma información a través de un gráfico que muestra la matriz de correlaciones mediante la librería `corrplot()`, donde los colores más oscuros demuestran mayores niveles de correlación.

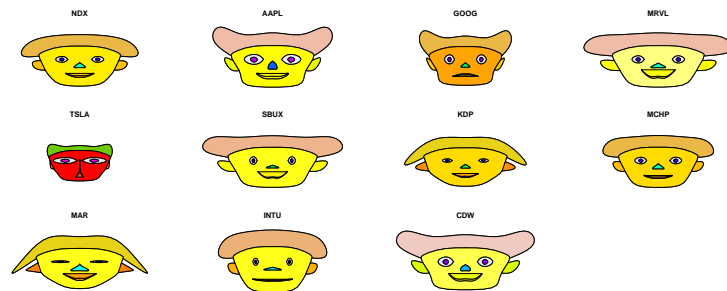
## 2. ANÁLISIS DE CORRELACIONES

---



### 3. Análisis descriptivo multivariante

Una vez realizado el análisis de correlaciones de los datos, puede resultar de interés plantear un análisis descriptivo multivariante que nos muestre, en qué medida, puede haber similitudes entre los datos analizados.



En esta línea, se puede recurrir a las caras de Chernoff, una técnica de visualización de datos planteada por Chernoff (1973), y basada en un *” método gráfico en el que ciertas características cuantitativas de un grupo, se asocian con datos físicos de la cara de una persona, con lo cual es posible realizar un dibujo que represente dichas características y realizar comparaciones”* (Fernández 2021).

Así, dicha técnica es reproducible a través de la librería `aplpack()`.

Podemos observar, a través de dichas caras de Chernoff y para los datos analizados, que las mayores similitudes se encuentran entre los valores cotizados AAPL,



COW, SBUX y MRVL. Por otra parte, KDP y MAR también podrían tener alguna similitud. La cara más discordante es la de la empresa norteamericana Tesla (TSLA), pues esto es debido a que es una empresa con unos rendimientos discordantes, especialmente en el cuatro trimestre en el que ha crecido más de un 50 %, mientras que el resto tuvo unos rendimientos más moderados.

## 4. Análisis de componentes

Una vez analizadas las correlaciones de los datos, y visualizados los datos mediante las caras de Chernoff, estamos en disposición de realizar un análisis de componentes principales (ACP), con el objetivo de determinar si una componente tiene la capacidad de explicar la mayor parte del comportamiento de los valores cotizados seleccionados en la muestra, y así, construir un índice nuevo. Es decir, en definitiva, el análisis ACP, nos permitirá transformar “*un conjunto de variables correlacionadas en un nuevo conjunto de variables incorrelacionadas*” (Fernández 2021).

### 4.1. Ejecución del análisis ACP

Veamos por tanto, los datos de las componentes principales para nuestro conjunto de datos. En el siguiente panel podemos observar que la PC1 recoge ya más del 60 % de la varianza y/o comportamiento de las variables originales recogidas por cada componente.

Importance of components:

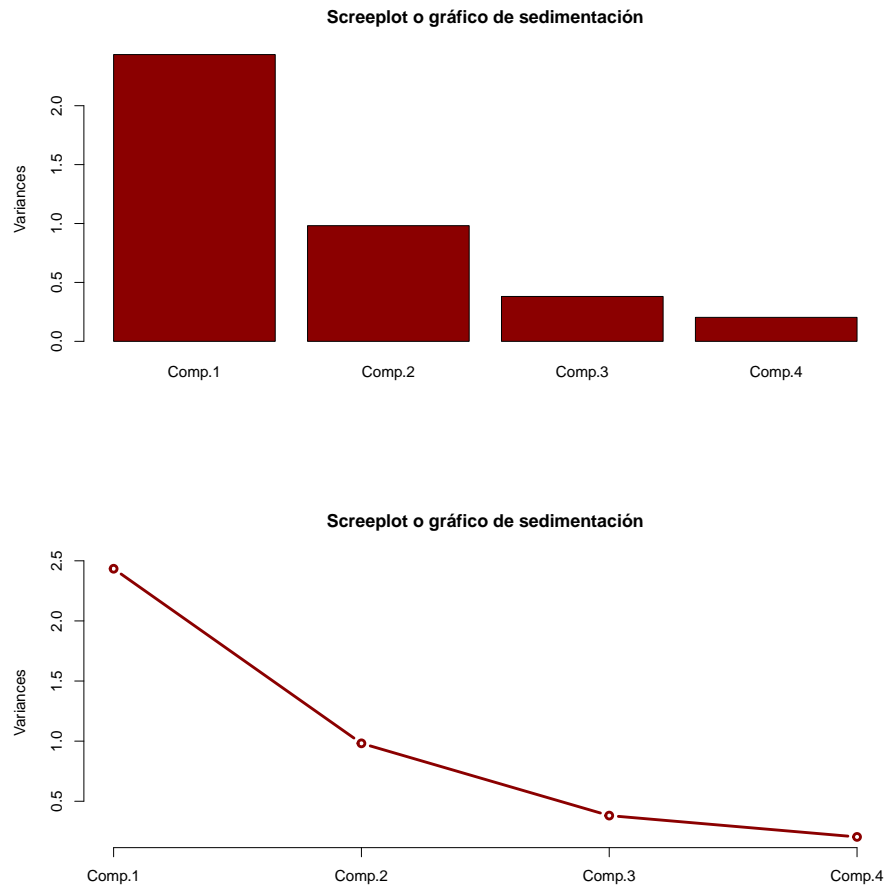
	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.5601356	0.9908134	0.61712017	0.45103046
Proportion of Variance	0.6085058	0.2454278	0.09520933	0.05085712
Cumulative Proportion	0.6085058	0.8539336	0.94914288	1.00000000

	Comp.1	Comp.2	Comp.3	Comp.4
T1	0.58	0.25	0.15	0.76
T2	0.55	-0.05	-0.79	-0.25
T3	-0.18	0.96	-0.14	-0.15
T4	-0.57	-0.10	-0.57	0.58

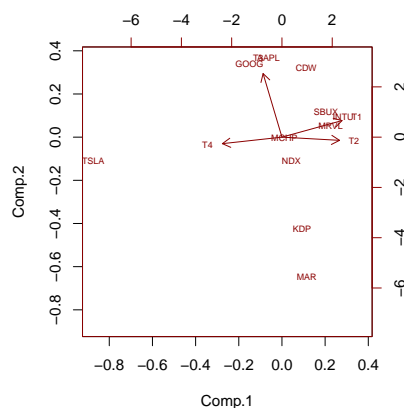
De esta forma, podemos representar los datos del ACP de forma gráfica mediante la librería `printcomp()`, la cual nos permite visualizar un *screeplot* o gráfico de sedimentación. En dicho gráfico vemos como la PC1, al igual que anteriormente comentábamos, parece recoger la mayor parte del comportamiento

#### 4.1. EJECUCIÓN DEL ANÁLISIS ACP 4. ANÁLISIS DE COMPONENTES

de las variables (rendimiento trimestral) del conjunto de datos de los valores cotizados.



En la misma línea, podemos observar un gráfico a partir de las puntuaciones del análisis ACP. Al mostrar gráficamente qué parte de la variación es recogida por cada componente, vemos que hay un salto significativo entre uno y dos componentes, pero sobre todo que el tercero y el cuarto explican bastante menos la varianza de nuestros datos. En los siguientes apartados indagaremos más en la cuestión y en sus detalles.



## 4.2. Componentes a retener

Una vez realizado en análisis de componentes principales, es fundamental determinar en qué medida se retienen más o menos componentes. Para ello existen distintos criterios que vamos a ver a continuación. Recordemos en este punto que el objetivo final es poder quedarnos con una única PC.

### 4.2.1. Criterio de varianza explicada acumulada

Este criterio considera retener las componentes en función de la varianza explicada acumulada según se seleccionen más o menos componentes. En este caso, siguiendo la recomendación de Morrison (1967), se deberían retener las dos primeras componentes (varianza explicada acumulada  $> 75\%$ ). En este caso, y teniendo en cuenta nuestro objetivo, desecharemos este criterio y optaremos por los resultados de los dos siguientes.

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.5601356	0.9908134	0.61712017	0.45103046
Proportion of Variance	0.6085058	0.2454278	0.09520933	0.05085712
Cumulative Proportion	0.6085058	0.8539336	0.94914288	1.00000000

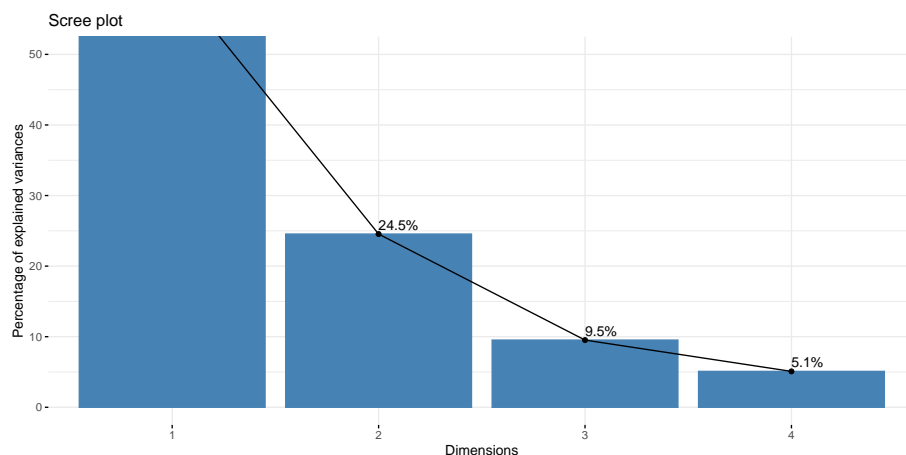
### 4.2.2. Criterio de la media aritmética (autovalores)

Este criterio considera retener tantas componentes como su valor, en la tabla de datos denominado como *eigenvalue*, sea superior a 1. En este caso, se confirma que con este criterio **retendríamos una única componente**.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.4340231	60.850578	60.85058
Dim.2	0.9817111	24.542777	85.39336
Dim.3	0.3808373	9.520933	94.91429
Dim.4	0.2034285	5.085712	100.00000

#### 4.2.3. Criterio de Cattell (gráfico de sedimentación)

Este criterio considera retener las componentes que se sitúan en la zona previa a la zona de sedimentación (cambio de tendencia de la curva). En este caso, al igual que en el anterior, también retendríamos **una única componente**.



#### 4.3. Interpretación de las componentes principales analizadas

En definitiva, se ha decidido retener una única componente (PC1) a partir del análisis de los distintos criterios de retención.

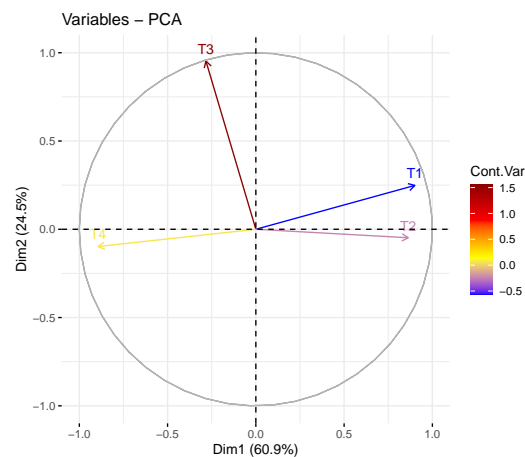
## 5. Presentación de resultados y conclusiones

En este último capítulo se procederá a presentar los resultados del análisis ACP y las conclusiones más relevantes. Así, para ello, se utilizará la función `get_pca_var()`, que contiene la siguiente información.

### Principal Component Analysis Results for variables

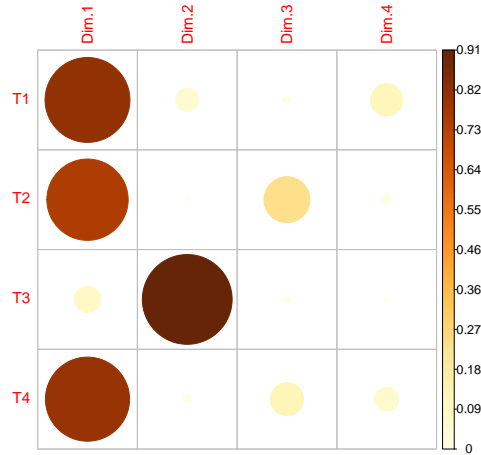
Name	Description
1 "\$coord"	"Coordinates for the variables"
2 "\$cor"	"Correlations between variables and dimensions"
3 "\$cos2"	"Cos2 for the variables"
4 "\$contrib"	"contributions of the variables"

En primer lugar, si observamos el círculo de correlación, vemos que los rendimientos del T1 y T2 son los que mayores correlaciones presentan entre sí positivamente. Por otra parte, la variable T3 sería la que mejor está representada en el mapa de factores en términos de calidad.

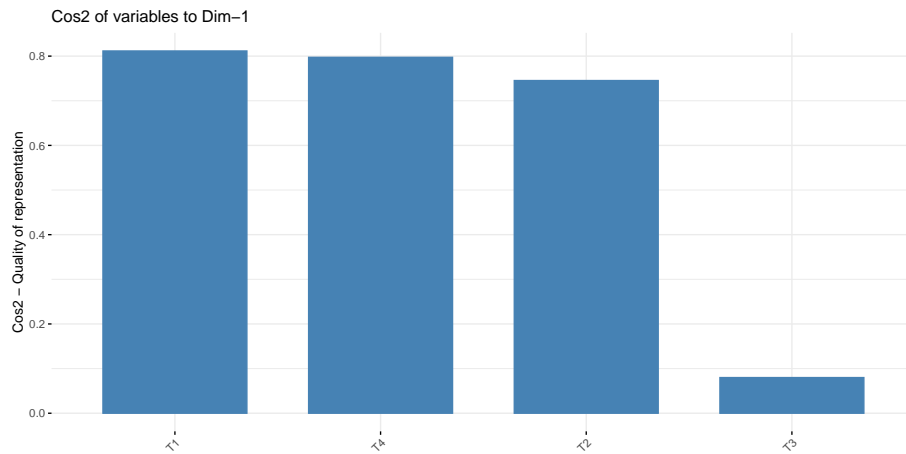


## 5. PRESENTACIÓN DE RESULTADOS Y CONCLUSIONES

En otro sentido, podemos plantear la calidad de las variables utilizadas (rendimientos) en las dimensiones. Así podemos ver que la mayor parte de variables están bien representadas en la PC1, a excepción de T3.



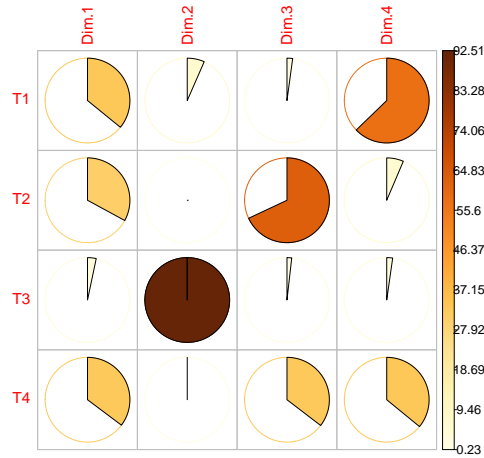
También es posible, en la misma línea, crear un diagrama de barras de las variables cos2 usando el paquete `factoextra()`, que nos muestra los mismos resultados que el anterior gráfico, donde T1 es la que mejor se representa en la PC1 a diferencia de T3 que es la que menos.



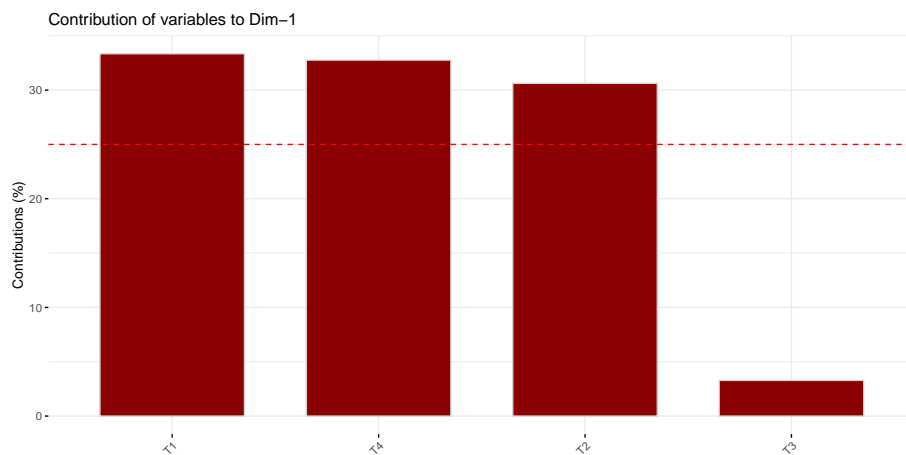
Podemos seguir analizando las variables y su relación con las componentes, determinando en qué medida contribuye cada variable a dichas componentes. En el siguiente gráfico se puede observar la contribución de las variables a las componentes, y podemos ver que a la componente primera, la retenida, contribuyen por igual forma T1, T2 y T4, mientras que T3 contribuye, especialmente, a la

## 5. PRESENTACIÓN DE RESULTADOS Y CONCLUSIONES

segunda componente, la cual si se retenía aumentaba casi un 20 % la explicación del comportamiento de las variables, aunque finalmente se había optado por el criterio de retener una única componente.

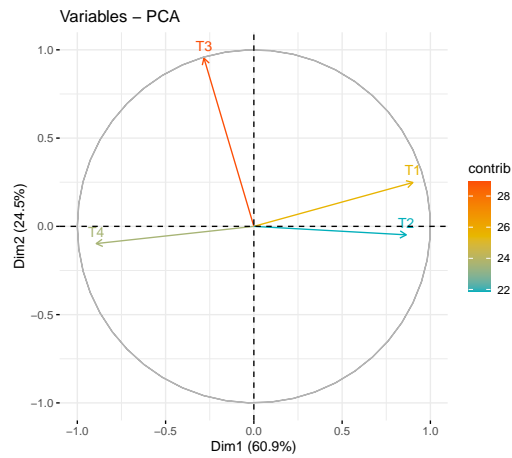


Esto mismo puede representarse mediante un gráfico de barras que muestra cuánto contribuye cada variable a la primera componente y otro gráfico *biplot* que resalta las variables más importantes en la contribución entre la PC1 y PC2.

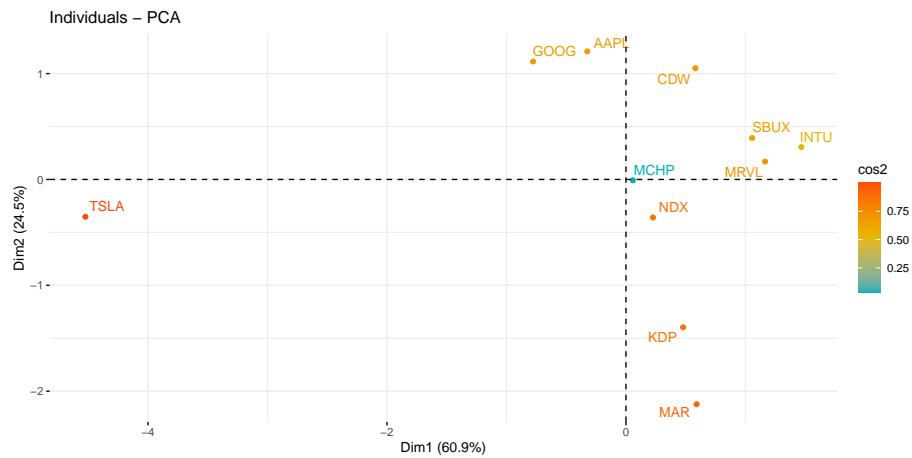




## 5. PRESENTACIÓN DE RESULTADOS Y CONCLUSIONES

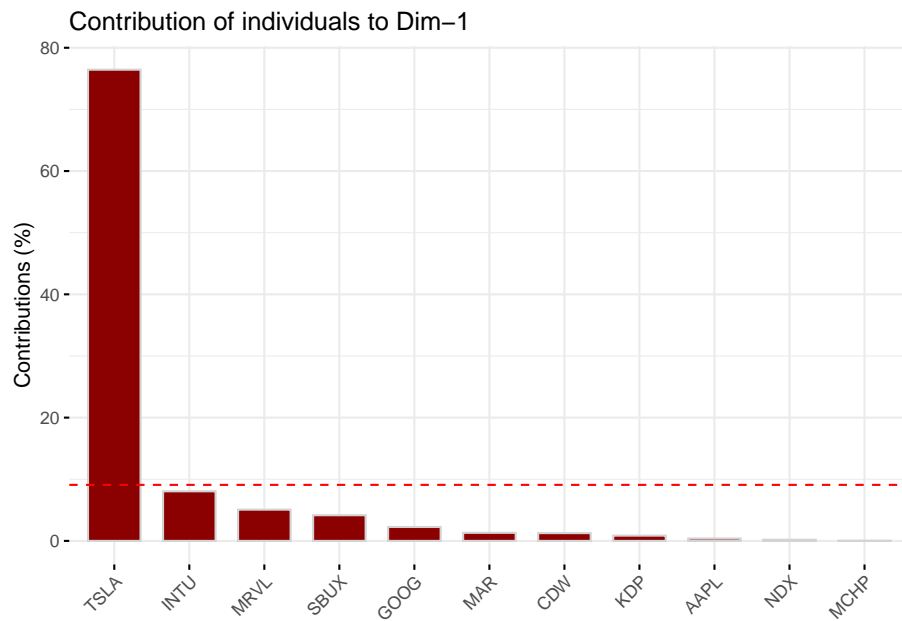


Siguiendo con el análisis, se pueden representar las observaciones (valores cotizados) y su representación en las dos primeras componentes (PC1 y PC2). Veamos como Tesla se aleja de forma significativa del área media donde se situán el resto de valores cotizados.



Por último, se puede observar que Tesla (TSLA) es el valor cotizado que más explica o contribuye a la primera componente retenida (PC1).

## 5. PRESENTACIÓN DE RESULTADOS Y CONCLUSIONES



En definitiva, las principales conclusiones que podemos extraer de este informe son:

- El análisis de componentes principales nos permite simplificar y mejorar la forma de obtener conclusiones.
- Se ha podido analizar los movimientos generales de precios de las empresas tecnológicas del NASDAQ en el 2019 a través de un único componente principal, ya sea separando los datos que tenemos por empresas como por trimestres.
- El tercer trimestre de 2019 se comporta de una manera distinta al resto del año.
- Tesla (TSLA) se comporta de una manera distinta al resto de empresas tecnológicas.

# Referencias bibliográficas

La gestión de referencias bibliográficas se ha realizado a través del gestor bibliográfico Mendeley® y posteriormente se ha integrado en el documento R Markdown con las ayudas proporcionadas por Ramos (2021), Van Hespén (2016) y Vidal (2011).

- Allaire, J. J., Rich Iannone, Alison P. Hill, y Yihui Xie. 2021. «Distill: R Markdown Format for Scientific and Technical Writing».
- AprendeR: Parte II. 2021. «AprendeR: Parte II». Tesis doctoral. <https://aprender-uib.github.io/AprendeR2/>.
- Cano, Emilio. 2021. «Introducción al software estadístico R». [https://www.lcano.com/b/iser/%7B/\\_%7Dbook/index.html](https://www.lcano.com/b/iser/%7B/_%7Dbook/index.html).
- Casero, Víctor M., y Consolación Quintana. 2021. «Tratamiento y Análisis Exploratorio de Datos Económicos». En *Máster Universitario en Modelización y Análisis de Datos Económicos (MUMADE)*.
- Chernoff, Herman. 1973. «The Use of Faces to Represent Points in K-Dimensional Space Graphically». *Journal of the American Statistical Association* 68 (342): 361-68. <https://doi.org/https://doi.org/10.1080/01621459.1973.10482434>.
- CRAN R-Project. 2021. «The YAML Fieldguide». <https://cran.r-project.org/web/packages/ymlthis/vignettes/yaml-fieldguide.html>.
- DataCamp. 2021. «RDocumentation». <https://www.rdocumentation.org/>.
- Fernández, Gema. 2021. «Técnicas de selección y síntesis de la información relevante: análisis de componentes principales. Aplicación con R.». En *Máster Universitario en Modelización y Análisis de Datos Económicos (MUMADE) - Técnicas Multivariantes Aplicadas al Análisis Sectorial - Universidad de Castilla-La Mancha*.
- Hlavac, Marek. 2018. «beautiful LaTeX, HTML and ASCII tables from R statistical output». <https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf>.
- Keyes, David. 2019. «How to make beautiful tables in R». <https://rfortherestofus.com/2019/11/how-to-make-beautiful-tables-in-r/>.
- Kobi. 2010. «Display a (R) or (TM) symbol on SE sites». <https://meta.stackexchange.com/a/68202>.
- Luque, Pedro L. 2019a. «Cómo crear tablas de información en R Markdown».

- Universidad de Sevilla. [http://destio.us.es/calvo/ficheros/ComoCrearTablasRMarkdown%7B/\\_%7DPedroLuque%7B/\\_%7D2019Sep%7B/\\_%7Dlibrodigital.pdf](http://destio.us.es/calvo/ficheros/ComoCrearTablasRMarkdown%7B/_%7DPedroLuque%7B/_%7D2019Sep%7B/_%7Dlibrodigital.pdf).
- . 2019b. «Construcción de tablas con knitr-kableExtra».
- Ramos, Marco. 2021. «Chapter 5. Integración Mendeley con R Markdown». [https://bookdown.org/marcoyel21/r%7B/\\_%7Dmarkdown%7B/\\_%7Dabogados/integraci%7B/'%7Bo%7D%7Dn-mendeley-con-r-markdown.html](https://bookdown.org/marcoyel21/r%7B/_%7Dmarkdown%7B/_%7Dabogados/integraci%7B/'%7Bo%7D%7Dn-mendeley-con-r-markdown.html).
- Tarancón, Miguel Á., y Gema Fernández. 2021. «Técnicas Multivariantes Aplicadas al Análisis Sectorial». En *Máster Universitario en Modelización y Análisis de Datos Económicos (MUMADE)*.
- Van Hespén, Rossana. 2016. «Writing your thesis with R Markdown (2) – Text, citations and equations». <https://rosannavanhespen.nl/rmarkdown/writing-your-thesis-with-r-markdown-2-text-citations-and-equations/>.
- Vidal, Ricardo. 2011. «HOWTO: Use Mendeley to create citations using LaTeX and BibTeX». <https://blog.mendeley.com/2011/10/25/howto-use-mendeley-to-create-citations-using-latex-and-bibtex/>.
- Xie, Yihui, J. J. Allaire, y Garrett Golemund. 2021. «R Markdown: The Definitive Guide». <https://bookdown.org/yihui/rmarkdown/>.
- Xie, Yihui, Christophe Dervieux, y Emily Riederer. 2021. «R Markdown Cookbook». <https://bookdown.org/yihui/rmarkdown-cookbook/>.
- Yahoo. 2021. «Yahoo! Finance». <https://es.finance.yahoo.com/>.
- Zhu, Hao. 2020. «Create awesome LaTeX table with knitr::kable and kableExtra». [https://haozhu233.github.io/kableExtra/awesome%7B/\\_%7Dtable%7B/\\_%7Din%7B/\\_%7Dpdf.pdf](https://haozhu233.github.io/kableExtra/awesome%7B/_%7Dtable%7B/_%7Din%7B/_%7Dpdf.pdf).

# Anexos

## Anexo 1. Datos de la sesión

En esta sección se recogen los datos de la sesión utilizada para elaborar este informe. Siguiendo a Cano (2021), es fundamental observar la versión de R, así como las versiones de los paquetes bajo los cuales se ha ejecutado el código o *script*.

```
R version 4.1.1 (2021-08-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19042)
```

```
Matrix products: default
```

```
locale:
```

```
[1] LC_COLLATE=Spanish_Spain.1252 LC_CTYPE=Spanish_Spain.1252
[3] LC_MONETARY=Spanish_Spain.1252 LC_NUMERIC=C
[5] LC_TIME=Spanish_Spain.1252
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
other attached packages:
```

```
[1] FactoMineR_2.4          factoextra_1.0.7
[3] PerformanceAnalytics_2.0.4 xts_0.12.1
[5] zoo_1.8-9               aplpack_1.3.5
[7] corrplot_0.92           ez_4.4-0
[9] GGally_2.1.2            ggplot2_3.3.5
[11] knitr_1.36              kableExtra_1.3.4
[13] dplyr_1.0.7             tidyr_1.1.4
[15] readxl_1.3.1
```

```
loaded via a namespace (and not attached):
```

```
[1] httr_1.4.2             viridisLite_0.4.0      splines_4.1.1
```

[4]	carData_3.0-4	assertthat_0.2.1	highr_0.9
[7]	cellranger_1.1.0	yaml_2.2.1	ggrepel_0.9.1
[10]	backports_1.3.0	pillar_1.6.4	lattice_0.20-45
[13]	glue_1.5.0	quadprog_1.5-8	digest_0.6.28
[16]	ggsignif_0.6.3	RColorBrewer_1.1-2	rvest_1.0.2
[19]	minqa_1.2.4	colorspace_2.0-2	htmltools_0.5.2
[22]	Matrix_1.3-4	plyr_1.8.6	pkgconfig_2.0.3
[25]	broom_0.7.10	purrr_0.3.4	scales_1.1.1
[28]	webshot_0.5.2	svglite_2.0.0	lme4_1.1-27.1
[31]	tibble_3.1.6	mgcv_1.8-38	farver_2.1.0
[34]	generics_0.1.1	car_3.0-12	ggpubr_0.4.0
[37]	DT_0.20	ellipsis_0.3.2	withr_2.4.2
[40]	magrittr_2.0.1	crayon_1.4.2	evaluate_0.14
[43]	fansi_0.5.0	nlme_3.1-153	MASS_7.3-54
[46]	rstatix_0.7.0	xml2_1.3.2	tools_4.1.1
[49]	lifecycle_1.0.1	stringr_1.4.0	munsell_0.5.0
[52]	cluster_2.1.2	flashClust_1.01-2	compiler_4.1.1
[55]	systemfonts_1.0.3	rlang_0.4.12	grid_4.1.1
[58]	nloptr_1.2.2.3	rstudioapi_0.13	htmlwidgets_1.5.4
[61]	leaps_3.1	labeling_0.4.2	tcltk_4.1.1
[64]	rmarkdown_2.11	boot_1.3-28	gtable_0.3.0
[67]	abind_1.4-5	DBI_1.1.1	reshape_0.8.8
[70]	reshape2_1.4.4	R6_2.5.1	fastmap_1.1.0
[73]	utf8_1.2.2	stringi_1.7.5	Rcpp_1.0.7
[76]	vctrs_0.3.8	scatterplot3d_0.3-41	tidyselect_1.1.1
[79]	xfun_0.28		

## Anexo 2. Base de datos

A continuación se presentan los datos utilizados e importados para elaborar el informe.

	<b>T1</b>	<b>T2</b>	<b>T3</b>	<b>T4</b>
NDX	0.148	0.025	-0.002	0.125
AAPL	0.189	0.038	0.109	0.264
GOOG	0.115	-0.100	0.105	0.103
MRVL	0.187	0.168	0.031	0.101
TSLA	-0.103	-0.258	0.059	0.528
SBUX	0.150	0.130	0.049	0.016
KDP	0.107	0.048	-0.071	0.071
MCHP	0.155	-0.004	0.021	0.126
MAR	0.155	0.100	-0.123	0.211
INTU	0.298	-0.015	0.008	-0.004
CDW	0.200	0.121	0.094	0.167

## Anexo 3. Código (*script*) utilizado

A continuación se presenta el *script* utilizado para desarrollar el informe.

```
[1] "---"
[2] "title: \"Análisis de componentes principales aplicado a un indicador de corporacione"
[3] "author:"
[4] "- PÃ©rez, R.S.^[Rafael Sergio PÃ©rez Pujol, UCLM, RafaelSergio.Perez@alu.uclm.es]"
[5] "- Bermann, M.A.^[Mateo Alberto Bermann Albalat, UCLM, MateoAlberto.Bermann@alu.uclm."
[6] "date: \"'r format(Sys.time(), '%d %B, %Y')\"'"
[7] "output:"
[8] "  pdf_document:"
[9] "    toc: yes"
[10] "  html_document:"
[11] "    toc: yes"
[12] "    df_print: paged"
[13] "header-includes:"
[14] "- \\usepackage{fancyhdr}"
[15] "- \\pagestyle{fancy}"
[16] "- \\fancyfoot[CO,CE]{TAEDE - MUMADE}"
[17] "- \\fancyfoot[LE,RO]{\\thepage}"
[18] "- \\usepackage{titling}"
[19] "- \\pretitle{\\begin{center} \\includegraphics[width=4in,height=4in]{logo_color.png}"
[20] "- \\posttitle{\\end{center}}}"
[21] "documentclass: report"
[22] "bibliography: library.bib"
[23] "lang: es"
[24] "---"
[25] ""
[26] "'{r, echo = FALSE, include = FALSE}"
[27] "# Aviso: si existen errores al compilar puede deberse a:"
[28] ""
[29] "# 1. Necesidad de incorporar el package \"tinytex\" [install.packages(\"tinytex\"); tin"
[30] "# 2. Necesidad de instalar LaTeX - MiKTeX (https://miktex.org/download)"
[31] "# 3. Posibles problemas de compilado: seguir pasos en https://yihui.org/tinytex/r/#d"
[32] ""
[33] ""
[34] "'{r setup, include=FALSE}"
[35] "# Ajustes iniciales de los chunk"
[36] "knitr::opts_chunk$set(echo = F, "
[37] "                        warning = F, "
[38] "                        message = F)"
[39] ""
[40] ""
[41] "'{r, include = FALSE}"
```



```

[42] "# Limpieza del entorno"
[43] "rm(list = ls())"
[44] ""
[45] "# Instalaci3n de paquetes no instalados"
[46] "packages <- c(\"readxl\", \"tidyr\", \"dplyr\", \"kableExtra\", \"
[47] \"          \"knitr\", \"ggplot2\", \"GGally\", \"ez\", \"
[48] \"          \"corrplot\", \"aplpack\", \"PerformanceAnalytics\", \"
[49] \"          \"factoextra\", \"FactoMineR\")"
[50] ""
[51] "installed_packages <- packages %in% rownames(installed.packages())"
[52] "if (any(installed_packages == FALSE)) {"
[53] "  install.packages(packages[!installed_packages])"
[54] "}"
[55] ""
[56] "# Activaci3n de paquetes"
[57] "library(readxl)"
[58] "library(tidyr)"
[59] "library(dplyr)"
[60] "library(kableExtra)"
[61] "library(knitr)"
[62] "library(ggplot2)"
[63] "library(GGally)"
[64] "library(ez)"
[65] "library(corrplot)"
[66] "library(aplpack)"
[67] "library(PerformanceAnalytics)"
[68] "library(factoextra)"
[69] "library(FactoMineR)"
[70] ""
[71] "# Importaci3n de datos"
[72] "datos_acp <- read_excel(\"TMAAS_3_database.xlsx\", \"
[73] \"          \"          sheet = \"database\")"
[74] "datos_acp <- data.frame(datos_acp, \"
[75] \"          row.names = 1)"
[76] "'''"
[77] ""
[78] "# Resumen"
[79] ""
[80] "Las grandes corporaciones tecnol3gicas han sufrido un crecimiento exponencial en el
, empresas como Apple, Google (Aplhabet) o Tesla han mejorado de forma considerable su posi
ndice NASDAQ 100 (~NDX) a partir de diferentes valores cotizados del propio Á-
ndice."
[81] ""
[82] "[^1]: Este informe ha sido realizado con el software R, a trav3s del entorno RStudio
culos y clases de @Allaire2021, @AprendeR2021, @Casero2021, @Cano2021, @CRANR-Project2021, @
[83] ""

```

```

[84] "# 1. Introducci3n"
[85] ""
[86] "A menudo, es f3cil perderse en un mar de informaci3n financiera, y a veces acabamos
ndice, en s3 mismo nos puede dar una ligera idea, pero sufre de varios problemas al ser un a
[87] ""
[88] "Es por ello por lo que, junto a esto, y en el marco del an3lisis ACP, el **objetivo
a la capacidad de explicar la mayor parte del comportamiento de los valores cotizados seleco
sucesivamente\", y donde el objetivo es \"ver si unas pocas componentes recogen la mayor p
, se puede arg3ir que la dimensionalidad del problema no es p sino inferior a p\"_."
[89] ""
[90] "Para conseguir el objetivo final se ha recurrido a la base de datos de @Yahoo2021, a
ndice NASDAQ 100. El per3odo de observaci3n incluye todo el a3o 2019 y se han obtenido los
[91] ""
[92] "[^2]: La fuente de la base de datos es @Yahoo2021."
[93] ""
[94] "En este punto, cabe decir que el an3lisis de componentes principales requiere que el
tulo."
[95] ""
[96] "Podemos ver de forma preliminar los datos de las 6 primeras observaciones (datos comp
[97] ""
[98] ""{r}"
[99] "datos_acp %>% "
[100] "  head() %>% "
[101] "  kable(booktabs = TRUE, "
[102] "        format = \"latex\", "
[103] "        caption = \"Vista inicial de los datos de las 6 primeras observaciones\", "
[104] "        digits = 1) %>% "
[105] "  kable_styling(font_size = 8, "
[106] "                latex_options = c(\"striped\", "
[107] "                                  \"condensed\", "
[108] "                                  \"hold_position\"), "
[109] "                position = \"center\", "
[110] "                full_width = F) %>% "
[111] "  row_spec(0, bold = T, color = \"black\")"
[112] ""{r}"
[113] ""
[114] "# 2. An3lisis de correlaciones"
[115] ""
[116] "Tal y como se ha comentado anteriormente, el an3lisis de componentes principales rec
, aplicando dicha premisa para el conjunto de datos utilizados para este informe, se puede c
[117] ""
[118] ""{r}"
[119] "# Matriz de correlaciones"
[120] "datos_acp %>% "
[121] "  cor() %>% "
[122] "  kable(booktabs = TRUE, "

```

```

[123] "          format = \"latex\", \"
[124] "          caption = \"Matriz de correlaciones valores e
ndice NASDAQ 100\", \"
[125] "          digits = 2) %>%\"
[126] "  kable_styling(font_size = 8, \"
[127] "          latex_options = c(\"striped\", \"
[128] "                          \"condensed\", \"
[129] "                          \"hold_position\"), \"
[130] "          position = \"center\", \"
[131] "          full_width = F) %>% \"
[132] "  row_spec(0, bold = T, color = \"black\")\"
[133] \"\"\"
[134] \"\"
[135] \"Así mismo, dicho análisis puede plantearse de forma gráfica a través de distintas
a 'ez()', se puede observar que dichos valores son relativamente elevados, especialmente par
[136] \"\"
[137] \"\"\"{r, fig.align = 'center', fig.width = 3, fig.height = 2}\"
[138] \"# Análisis de correlaciones con la librería ez()\"
[139] \"ezCor(datos_acp)\"
[140] \"\"\"
[141] \"\"
[142] \"De la misma forma, podemos observar la misma información a través de un gráfico qu
a 'corrplot()', donde los colores más oscuros demuestran mayores niveles de correlación.\"
[143] \"\"
[144] \"\"\"{r, fig.align = 'center', fig.width = 3, fig.height = 3}\"
[145] \"# Análisis de correlaciones con la librería corrplot()\"
[146] \"corrplot(cor(datos_acp), method=\"number\", number.cex = .7)\"
[147] \"\"\"
[148] \"\"
[149] \"# 3. Análisis descriptivo multivariante\"
[150] \"\"
[151] \"Una vez realizado el análisis de correlaciones de los datos, puede resultar de inter
[152] \"\"
[153] \"\"\"{r, fig.align = 'center', fig.width = 10, fig.height = 6}\"
[154] \"# Visualización de datos con Chernoff\"
[155] \"faces(datos_acp, \"
[156] \"      cex = 0.8, \"
[157] \"      print.info = F, \"
[158] \"      face.type = 1)\"
[159] \"\"\"
[160] \"\"
[161] \"En esta línea, se puede recurrir a las caras de Chernoff, una técnica de visualizaci
sticas cuantitativas de un grupo, se asocian con datos físicos de la cara de una persona, co
sticas y realizar comparaciones\"_ [Fernandez2021a]. \"
[162] \"\"
[163] \"Así, dicha técnica es reproducible a través de la librería-

```

```

a 'aplpack()'.
[164] ""
[165] "Podemos observar, a través de dichas caras de Chernoff y para los datos analizados,
an tener alguna similitud. La cara más discordante es la de la empresa norteamericana Tesla
[166] ""
[167] "# 4. Análisis de componentes"
[168] ""
[169] "Una vez analizadas las correlaciones de los datos, y visualizados los datos mediante
, construir un Índice nuevo. Es decir, en definitiva, el análisis ACP, nos permitirán trans
[170] ""
[171] "## 4.1. Ejecución del análisis ACP"
[172] ""
[173] "Veamos por tanto, los datos de las componentes principales para nuestro conjunto de c
[174] ""
[175] "```{r, comment = ''}"
[176] "fit1 <- princomp(datos_acp, "
[177] "                cor = TRUE)"
[178] "fit1 %>% "
[179] "  summary()"
[180] "```"
[181] ""
[182] "```{r, comment = ''}"
[183] "pc1 <- unclass(loadings(fit1))"
[184] "round(pc1, "
[185] "      digits = 2)"
[186] "```"
[187] ""
[188] "De esta forma, podemos representar los datos del ACP de forma gráfica mediante la l
a 'printcomp()', la cual nos permite visualizar un _screeplot_ o gráfico de sedimentación.
[189] ""
[190] "```{r, fig.align = 'center', fig.width = 10, fig.height = 5}"
[191] "screeplot(fit1, "
[192] "          col = \"red4\", "
[193] "          main = \"Screeplot o gráfico de sedimentación\")"
[194] "```"
[195] ""
[196] "```{r, fig.align = 'center', fig.width = 10, fig.height = 5}"
[197] "plot(fit1,"
[198] "     type = \"lines\", "
[199] "     col = \"red4\", "
[200] "     lwd = \"3\", "
[201] "     main = \"Screeplot o gráfico de sedimentación\")"
[202] "```"
[203] ""
[204] "En la misma línea, podemos observar un gráfico a partir de las puntuaciones del análisis
[205] ""

```

```

[206] "{'r, fig.align = 'center', fig.width = 10, fig.height = 5}"
[207] "pc.datos_acp <- fit1$scores"
[208] "biplot(fit1, "
[209] "      cex = 0.6,"
[210] "      col = 'red4')"
[211] "'''"
[212] ""
[213] "## 4.2. Componentes a retener"
[214] ""
[215] "Una vez realizado en análisis de componentes principales, es fundamental determinar
[216] ""
[217] "### 4.2.1. Criterio de varianza explicada acumulada"
[218] ""
[219] "Este criterio considera retener las componentes en función de la varianza explicada
an retener las dos primeras componentes (varianza explicada acumulada > 75%). En este caso,
[220] ""
[221] "{'r, comment = ''}"
[222] "fit1 %>% "
[223] "  summary()"
[224] "'''"
[225] ""
[226] "### 4.2.2. Criterio de la media aritmética (autovalores)"
[227] ""
[228] "Este criterio considera retener tantas componentes como su valor, en la tabla de datos
amos una única componente**."
[229] ""
[230] "{'r, comment = ''}"
[231] "# Criterio media aritmética (autovalores)"
[232] "res.pca <- PCA(datos_acp, "
[233] "      graph = FALSE)"
[234] "eig.val <- get_eigenvalue(res.pca)"
[235] "eig.val"
[236] "'''"
[237] ""
[238] "### 4.2.3. Criterio de Cattell (gráfico de sedimentación)"
[239] ""
[240] "Este criterio considera retener las componentes que se sitúan en la zona previa a la
amos **una única componente**."
[241] ""
[242] "{'r, fig.align = 'center', fig.width = 10, fig.height = 5}"
[243] "# Criterio de Cattell (gráfico sedimentación)"
[244] "fviz_eig(res.pca, "
[245] "      addlabels = TRUE, "
[246] "      ylim = c(0, 50))"
[247] "'''"
[248] ""

```

```

[249] "## 4.3. Interpretaci3n de las componentes principales analizadas"
[250] ""
[251] "En definitiva, se ha decidido retener una 1nica componente (PC1) a partir del an1lisis ACP y
[252] ""
[253] "# 5. Presentaci3n de resultados y conclusiones"
[254] ""
[255] "En este 1ltimo cap1tulo se proceder1 a presentar los resultados del an1lisis ACP y
, para ello, se utilizar1 la funci3n 'get_pca_var()', que contiene la siguiente informaci3n
[256] ""
[257] "```{r, comment = ''}"
[258] "# Informaci3n contenida"
[259] "var <- get_pca_var(res.pca)"
[260] "var"
[261] "```"
[262] ""
[263] "En primer lugar, si observamos el c1rculo de correlaci3n, vemos que los rendimientos
positivamente. Por otra parte, la variable T3 ser1a la que mejor est1 representada en el ma
[264] ""
[265] "```{r, fig.align = 'center', fig.width = 10, fig.height = 5}"
[266] "# C1rculo de correlaci3n"
[267] "set.seed(123)"
[268] "my.cont.var <- rnorm(4)"
[269] ""
[270] "# Color variables"
[271] "fviz_pca_var(res.pca, "
[272] "             col.var = my.cont.var,"
[273] "             gradient.cols = c(\"blue\", \"yellow\", \"red\", \"red4\"),\"
[274] "             legend.title = \"Cont.Var\")"
[275] "```"
[276] ""
[277] "En otro sentido, podemos plantear la calidad de las variables utilizadas (rendimiento)
podemos ver que la mayor parte de variables est1n bien representadas en la PC1, a excepci3n
[278] ""
[279] "```{r, fig.align = 'center', fig.width = 10, fig.height = 5}"
[280] "# Representaci3n de las variables en las dimensiones"
[281] "corrplot(var$cos2, "
[282] "          is.corr = FALSE)"
[283] "```"
[284] ""
[285] "Tambi3n es posible, en la misma l1nea, crear un diagrama de barras de las variables
[286] ""
[287] "```{r, fig.align = 'center', fig.width = 10, fig.height = 5}"
[288] "# Gr1fico barras cos2"
[289] "fviz_cos2(res.pca, "
[290] "           choice = \"var\", "
[291] "           axes = 1)"

```

```

[292] ""
[293] ""
[294] "Podemos seguir analizando las variables y su relación con las componentes, determinando
aumentaba casi un 20% la explicación del comportamiento de las variables, aunque finalmente
a optado por el criterio de retener una única componente."
[295] ""
[296] ""
[297] "corrplot(var$contrib, "
[298] "          method = \"pie\", "
[299] "          is.corr = FALSE)"
[300] ""
[301] ""
[302] "Esto mismo puede representarse mediante un gráfico de barras que muestra cuánto con
[303] ""
[304] ""
[305] "par(mfrow=c(1,2))"
[306] "fviz_contrib(res.pca, "
[307] "             choice = \"var\", "
[308] "             fill = \"red4\", "
[309] "             color = \"lightgray\", "
[310] "             axes = 1, "
[311] "             top = 10)"
[312] ""
[313] ""
[314] ""
[315] ""
[316] "fviz_pca_var(res.pca, "
[317] "             col.var = \"contrib\", "
[318] "             gradient.cols = c(\"#00AFBB\", "
[319] "                               \"#E7B800\", "
[320] "                               \"#FC4E07\"))"
[321] ""
[322] ""
[323] "Siguiendo con el análisis, se pueden representar las observaciones (valores cotizados)
[324] ""
[325] ""
[326] "fviz_pca_ind(res.pca, col.ind = \"cos2\", "
[327] "             gradient.cols = c(\"#00AFBB\", "
[328] "                               \"#E7B800\", "
[329] "                               \"#FC4E07\"), "
[330] "             repel = TRUE)"
[331] ""
[332] ""
[333] "Por último, se puede observar que Tesla (TSLA) es el valor cotizado que más explica
[334] ""
[335] ""

```

```

[336] "fviz_contrib(res.pca, "
[337] "             choice = \"ind\", "
[338] "             axes = 1, "
[339] "             fill = \"red4\", "
[340] "             color = \"lightgray\")"
[341] ""
[342] ""
[343] ""
[344] "En definitiva, las principales conclusiones que podemos extraer de este informe son:"
[345] ""
[346] " + El análisis de componentes principales nos permite simplificar y mejorar la forma
[347] " + Se ha podido analizar los movimientos generales de precios de las empresas tecnológicas
[348] " + El tercer trimestre de 2019 se comporta de una manera distinta al resto del año.
[349] " + Tesla (TSLA) se comporta de una manera distinta al resto de empresas tecnológicas
[350] " "
[351] "# Referencias bibliográficas"
[352] ""
[353] "> La gestión de referencias bibliográficas se ha realizado a través del gestor bibliográfico
[354] ""
[355] "<div id=\"refs\"></div>"
[356] ""
[357] "# Anexos"
[358] ""
[359] "## Anexo 1. Datos de la sesión"
[360] ""
[361] "En esta sección se recogen los datos de la sesión utilizada para elaborar este informe,
como las versiones de los paquetes bajo los cuales se ha ejecutado el código o _script_."
[362] ""
[363] ""
[364] "sessionInfo()"
[365] ""
[366] ""
[367] "\\newpage"
[368] ""
[369] "## Anexo 2. Base de datos"
[370] ""
[371] "A continuación se presentan los datos utilizados e importados para elaborar el informe
[372] ""
[373] ""
[374] "datos_acp %>% "
[375] "  kable(booktabs = TRUE, "
[376] "        format = \"latex\", "
[377] "        digits = 3) %>% "
[378] "  kable_styling(font_size = 8, "
[379] "               latex_options = c(\"striped\", "
[380] "                                \"condensed\", "

```



```
[381] "                                \"hold_position\"), \"
[382] "                                position = \"center\", \"
[383] "                                full_width = F) %>% \"
[384] \" row_spec(0, bold = T, \"
[385] \"                                color = \"black\")\"
[386] \"\"\"
[387] \"\"
[388] \"\\newpage\"
[389] \"\"
[390] \"## Anexo 3. Código (_script_) utilizado\"
[391] \"\"
[392] \"A continuación se presenta el _script_ utilizado para desarrollar el informe.\"
[393] \"\"
[394] \"\"\"{r, echo = FALSE, comment= ''}\"
[395] \"script <- readLines(\"TMAAS_03.Rmd\")\"
[396] \"print(script)\"
[397] \"\"\"
```