

Introducción a los SIG con R



Dominic Royé
Roberto Serrano Notivoli

Introducción a los SIG con R

Contenido

R tiene, como lenguaje de programación enfocado al análisis estadístico, todos los ingredientes para ser usado como herramienta de análisis espacial y representación cartográfica: es gratuito, permite personalizar, replicar y compartir los análisis de cualquier nivel de dificultad y carece de limitaciones en cuanto a cantidad de información a procesar o tipos de formato diferentes para gestionar.

Todo ello lo coloca en una situación de ventaja que mejora día a día, gracias a su amplia comunidad de usuarios, respecto a un SIG (Sistema de Información Geográfica) convencional. Este manual explica, sin necesidad de conocimientos previos, cómo desarrollar con R todos los análisis disponibles en un SIG, con ejemplos sencillos y multitud de casos prácticos. Además, se muestran las enormes posibilidades de representación cartográfica, que van mucho más allá de la simple creación de mapas. R permite desde exportar a cualquier formato de archivo hasta crear mapas dinámicos para su publicación en Internet.

DOMINIC ROYÉ (Grevenbroich, Alemania, 1981)

Doctor en Geografía Física e investigador postdoctoral en la Universidad de Santiago de Compostela. Es especialista en análisis espacial y visualización de datos, con amplia experiencia en investigación y docencia en Biometeorología y Geografía Física en centros de investigación de España y Portugal.

ROBERTO SERRANO NOTIVOLI (Zaragoza, España, 1983)

Doctor en Geografía e investigador postdoctoral en el Consejo Superior de Investigaciones Científicas. Su investigación en modelización climática y sus aplicaciones se ha plasmado en más de treinta publicaciones y proyectos nacionales e internacionales.

Introducción
a los SIG con R



Dominic Royé
Roberto Serrano Notivoli

PRENSAS DE LA UNIVERSIDAD DE ZARAGOZA

COLECCIÓN: **Textos docentes, 286**

TEMÁTICA: **Geografía**

AÑO EDICIÓN: **2019**

Nº PÁGINAS: **358 pp**

FORMATO: **16 x 23 cm**

ENCUADERNACIÓN: **Rústica**

IDIOMA: **Castellano**

ISBN: **978-84-1340-002-0**

PVP: **36 €**



C/ Pedro Cerbuna, 12
50009 Zaragoza. España
Tel. 976 553 156
puz@unizar.es
puz.unizar.es



Prensas de la Universidad
Universidad Zaragoza

Presentación

Este libro nace de nuestra experiencia en el día a día con R a lo largo de la última década, pero, sobre todo, de una necesidad en el mundo de la cartografía y el análisis espacial: personalizar al máximo los análisis haciéndolos eficientes, fiables y reproducibles. En la era de los conjuntos de datos inacabables, la inmediatez de los resultados y el intercambio de información, R es el lenguaje perfecto para comunicarse: es gratuito, permite explorar hasta el último rincón de su funcionamiento y, por encima de todo, tiene una comunidad de usuarios enorme que no para de crecer y compartir soluciones, ideas y código, mucho código...

En los últimos años, a través de muchos cursos de formación en lenguaje R, hemos observado que, más allá de la evidente necesidad de una formación sólida en análisis estadístico, el usuario de Sistemas de Información Geográfica (SIG) que comienza en R tiene unas demandas muy claras. Por una parte busca unos conocimientos básicos con los que comenzar en el mundo de la programación (si es que no tenía ya experiencia previa) y, por otra, quiere saber cómo hacer lo que ya sabe, pero de una manera más eficiente y recursiva.

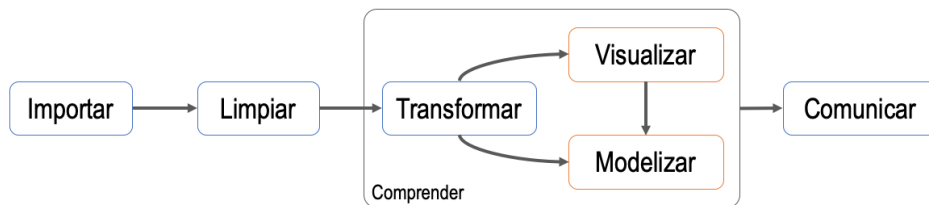
Los SIG son herramientas para la manipulación de datos espaciales con magníficas funciones de visualización. Sin embargo, la mayor parte del *software* disponible en el mercado, excepto contadas (y muy buenas) excepciones, tiene al menos dos inconvenientes: 1) no es gratuito y 2) es muy potente para el análisis espacial cuando hay que hacerlo unas pocas veces, pero se vuelve ineficiente cuando hay que repetirlo cientos o miles de veces. Lo que sí es común a todas estas herramientas es que ninguna es capaz de manejar enormes volúmenes de información. Hablamos de *Big Data*, lo cual no hace referencia exclusivamente a grandes cantidades de datos, sino a ser capaces de extraer resultados, patrones y/o informaciones útiles de ellos. Los SIG, en general, tienen serias limitaciones a este respecto, bien sea por una gestión ineficiente de los recursos, que suelen estar dedicados a una interfaz gráfica potente, bien sea por una simple falta de funciones de análisis adecuadas.

R se presenta como una forma de solucionar estos problemas. Si bien no es el único sistema con estas posibilidades, ya que existen muchos otros lenguajes de programación que son capaces de hacerlo, se trata de uno de los más completos. Todo esto ha sido posible, en los últimos años, gracias a los usuarios que han desarrollado herramientas para replicar todas y cada una de las funciones de un SIG, pero con las capacidades que, por su naturaleza, tiene el lenguaje R.

Flujo de trabajo en R y análisis espacial

Como todos los lenguajes de programación y a diferencia de los SIG, R tiene unas normas de funcionamiento básicas (ver capítulo de Introducción) para leer archivos, operar con los datos, hacer transformaciones o exportar los resultados. Pero aparte de esto, que no son más que una serie de reglas que hay que conocer, el flujo de trabajo ideal con R no dista mucho del de un SIG.

La siguiente figura muestra una conceptualización de este flujo, cuyas fases veremos en detalle en los capítulos sucesivos. Mientras algunas de ellas pueden obviarse en función del tipo de información que estamos trabajando o del tipo de análisis que queramos hacer, hay otras que son inevitables: importación y comunicación son básicas. La primera requiere de un conocimiento técnico (poco profundo en R) para leer la información, y para la segunda necesitamos tener claras al menos dos ideas: 1) en qué formato están los resultados y 2) a qué público van dirigidos. Esto nos permitirá ser más eficientes en nuestros análisis. Por otro lado, entre la importación y la comunicación hay una fase de comprensión que es donde realmente utilizamos todo el potencial de R. Es aquí donde centraremos los esfuerzos de este libro y a los que dedicaremos la mayor parte de los capítulos intentando mostrar, con ejemplos, muchas de las posibilidades del lenguaje R en análisis espacial y cartografía.



Flujo de trabajo en R. Fuente: Adaptado de Grolemund & Wickham: *R for Data Science*, 2017.

¿A quién está dirigido este libro?

Este libro es para todo aquel que esté interesado en el análisis espacial y la cartografía de unos pocos datos o de bases de datos ilimitadas. Quizá ya sabes algo o eres experto/a en SIG y esta es tu primera toma de contacto con R, o quizá ya conoces el lenguaje de programación y quieres saber cómo hacer todo lo que ya haces con los SIG en R. En cualquier caso, este manual puede ayudarte a mejorar

tus aptitudes en análisis espacial y representación cartográfica con R. Sería deseable que tuvieses alguna noción previa básica sobre cartografía temática, proyecciones o representación gráfica en general. No obstante, no es en absoluto necesario, ya que hemos intentado detallar los procedimientos, sin entrar en explicaciones profundas, para que la lectura sea más agradable.

¿Qué contiene este libro?

El libro está dividido en 9 capítulos:

Capítulo 1. Introducción: Breve introducción al uso de R para importar datos, hacer algunas transformaciones, conocer los tipos y clases de objetos y exportar los resultados.

Capítulo 2. Tidyverse: Explicación de los procedimientos posibles con este completísimo y actual paquete de funciones que nos permitirá importar, manipular y gestionar información alfanumérica de una manera sencilla y eficiente.

Capítulo 3. ggplot2: Representación gráfica de datos con uno de los paquetes de funciones más populares del momento en R. Se da un repaso a los distintos tipos de representación posibles, así como a diversas opciones de visualización.

Capítulo 4. Consideraciones iniciales: Breve explicación de algunos conceptos básicos sobre análisis espacial y cartografía que pueden ser útiles en caso de que aún no estés familiarizado con ellos.

Capítulo 5. Análisis vectorial: Repaso de las funciones de análisis que pueden aplicarse sobre datos espaciales de naturaleza vectorial. Introducción al uso de diferentes paquetes de funciones.

Capítulo 6. Análisis ráster: Similar al capítulo anterior, pero dedicado a los datos espaciales de naturaleza ráster.

Capítulo 7. Visualización de mapas: Explicación de los diversos modos de visualización cartográfica, estática e interactiva.

Capítulo 8. Casos prácticos: Colección de ejemplos en los que se pone en práctica la mayor parte de las funciones de análisis espacial estudiadas en los capítulos anteriores.

Capítulo 9. Extras: Introducción a algunas funciones o métodos adicionales de R que pueden ser de utilidad en el análisis espacial y la representación cartográfica.

Dónde encontrar los datos

Todos los datos utilizados en los ejemplos de cada capítulo están disponibles en el repositorio de *GitHub* accesible a través del siguiente enlace: https://github.com/dominicroye/libro_sig_r. El *feedback* de los lectores es siempre bienvenido, tanto del libro en general como de los datos en particular. En el mismo repositorio de los datos existe una sección (*Issues*) en la que es posible interactuar con los autores a este respecto.

Agradecimientos

Estamos enormemente agradecidos a toda la comunidad de usuarios de R que aportan soluciones y novedades desinteresadamente, en especial a través de los nuevos paquetes de funciones en el repositorio oficial (<https://cran.r-project.org>) y a través de la resolución de problemas en *stackoverflow* (<https://stackoverflow.com/questions/tagged/r>).

Queremos agradecer en particular, a Santiago Beguería su experiencia, consejos, correcciones y, en definitiva, su ayuda inestimable para escribir este libro. También a los organizadores de los cursos de formación de estos últimos años: José María Cuadrat, Alberto Martí, Ramón Chao y Miguel Ángel Saz. Por supuesto, también a todos los alumnos que han pasado por la experiencia de enfrentarse a R por primera vez. De ellos surge la idea y necesidad de escribir este manual. Gracias.

Los autores

Dominic Royé (Grevenbroich, Alemania, 1981) es licenciado en Geografía y Filología Hispánica por la Universidad de Colonia y la Universidad Técnica de Aquisgrán (Alemania). En 2015 obtuvo el doctorado en Geografía Física por la Universidad de Santiago de Compostela y actualmente es investigador postdoctoral en ella. Sus principales líneas de investigación son, por una parte, la Biometeorología y la Geografía de la salud, la relación entre la salud humana y el ambiente atmosférico, y, por otra, la Geografía física aplicada centrada en las variables atmosféricas y su comportamiento espacio-temporal. Se encuentra sumergido en el mundo de R desde el año 2010 y es un entusiasta usuario con mucha curiosidad por el análisis espacial, los SIG y la visualización, gestión y

manipulación de datos. Ha sido docente en varios cursos de lenguaje R en la Universidad de Barcelona, Universidade do Porto, Universidad Politécnica de Barcelona y Universidad de Santiago de Compostela.

Contacto: <https://dominicroye.github.io/en/>

Roberto Serrano Notivoli (Zaragoza, España, 1983) es licenciado y doctor en Geografía y Ordenación del Territorio por la Universidad de Zaragoza (España), especializado en climatología, Sistemas de Información Geográfica y cartografía. Actualmente es investigador postdoctoral Juan de la Cierva en la Estación Experimental de Aula Dei (Consejo Superior de Investigaciones Científicas) en Zaragoza. Su investigación está centrada principalmente en la modelización climática instrumental y sus aplicaciones en hidrología, ecología, reconstrucción paleoclimática, agroclimatología o riesgos ambientales, entre otras. Con una década de experiencia en R, está interesado en el lenguaje a nivel de desarrollo y en códigos eficientes para supercomputación y gestión y manipulación de conjuntos de datos masivos. Tiene un amplio bagaje docente en cursos de R en España y Chile, así como experiencia en el desarrollo de paquetes de funciones.

Este libro se escribió usando la versión de R 3.6.1 y de RStudio 1.2.1.

Índice

| | |
|--------------------|---|
| PRESENTACIÓN | 7 |
|--------------------|---|

PARTE 1. INTRODUCCIÓN A R

| | |
|-----------------------------------------------------------------------|----|
| 1. BREVE INTRODUCCIÓN AL LENGUAJE R..... | 15 |
| 1.1. ¿Qué es R y por qué debemos aprender a usarlo? | 15 |
| 1.2. Instalación de R e interfaz de RStudio | 16 |
| 1.3. Crear proyectos..... | 18 |
| 1.4. Primeros pasos: operaciones básicas | 19 |
| 1.5. Tipo de datos y operaciones aritméticas..... | 20 |
| 1.6. Objetos: modos y clases | 22 |
| 1.6.1. Vectores..... | 23 |
| 1.6.2. Factores | 26 |
| 1.6.3. Matrices..... | 26 |
| 1.6.4. Data.frames | 29 |
| 1.6.5. Series temporales | 30 |
| 1.6.6. Listas | 31 |
| 1.6.7. Conversión entre objetos..... | 32 |
| 1.6.8. Resumen de clases de objetos | 33 |
| 1.7. Funciones..... | 33 |
| 1.7.1. Ayuda sobre las funciones..... | 35 |
| 1.7.2. Paquetes de funciones | 35 |
| 1.8. Conceptos básicos sobre lectura, escritura y almacenamiento | 36 |
| 1.9. Algunas funciones de estadística fundamental..... | 38 |
| 1.9.1. Promedio, mediana, varianza | 39 |
| 1.9.2. Test..... | 40 |
| 1.9.3. Regresión lineal..... | 42 |
| 1.10. Trabajando con <i>scripts</i> | 44 |
| 1.11. Algunos mensajes de error comunes | 45 |
| 1.12. Consejos de estilo en programación con R | 46 |
| 2. MANIPULACIÓN Y GESTIÓN DE DATOS CON <i>TIDYVERSE</i> | 49 |
| 2.1. Introducción a la colección <i>tidyverse</i> | 49 |

| | |
|-------------------------------------------------------|-----|
| 2.2. Lectura y escritura | 52 |
| 2.3. Manipulación de caracteres | 56 |
| 2.4. Manipulación de fechas y horas | 60 |
| 2.5. Manipulación de tablas y vectores | 67 |
| 2.6. Otras funciones útiles | 81 |
| 2.7. Importar varias hojas desde Excel | 84 |
| 3. VISUALIZACIÓN DE DATOS CON <i>GGPLOT2</i> | 93 |
| 3.1. Introducción..... | 93 |
| 3.2. La gramática de gráficos en <i>ggplot2</i> | 97 |
| 3.2.1. Primeros pasos | 97 |
| 3.2.2. Gráficos de puntos..... | 101 |
| 3.2.3. Cambiar colores | 106 |
| 3.2.4. Transformaciones estadísticas..... | 111 |
| 3.2.5. Cambios en los ejes..... | 112 |
| 3.2.6. Cambiar el estilo del gráfico | 114 |
| 3.2.7. Gráficos de facetas | 116 |
| 3.2.8. Histograma y gráfico de densidad..... | 123 |
| 3.2.9. Gráficos de líneas..... | 128 |
| 3.2.10. <i>Boxplot</i> | 132 |
| 3.2.11. Gráficos de barras..... | 133 |
| 3.2.12. Gráficos de áreas | 141 |
| 3.2.13. <i>Heatmaps</i> | 143 |
| 3.3. Opciones adicionales..... | 145 |
| 3.3.1. Etiquetas de texto superpuestas..... | 145 |
| 3.3.2. Mosaico de gráficos | 146 |
| 3.3.3. Exportación de gráficos..... | 148 |

PARTE 2.

INTRODUCCIÓN AL ANÁLISIS ESPACIAL CON R

| | |
|-----------------------------------------------------------|-----|
| 4. CONSIDERACIONES INICIALES SOBRE ANÁLISIS ESPACIAL..... | 153 |
| 4.1. Formatos de datos..... | 153 |
| 4.2. Proyecciones..... | 156 |
| 4.2.1. Sistemas de coordenadas geográficos | 156 |
| 4.2.2. Sistemas de coordenadas proyectados..... | 157 |
| 4.2.3. Transformaciones | 157 |
| 4.3. Exportación de datos | 158 |

| | | |
|----------|---------------------------------------------------------------------------------|-----|
| 5. | DATOS ESPACIO-TEMPORALES I: VECTORIAL | 159 |
| 5.1. | Los paquetes <i>sp</i> y <i>rgdal</i> | 160 |
| 5.1.1. | Importación | 161 |
| 5.1.2. | Visualización | 162 |
| 5.1.3. | Proyecciones y coordenadas | 163 |
| 5.1.4. | Acceso a las variables | 167 |
| 5.1.5. | Funciones de análisis espacial | 168 |
| 5.1.5.1. | Unir tabla (<i>Join</i>) | 168 |
| 5.1.5.2. | Filtrado (<i>Select by attributes</i>) y Unión (<i>Append</i>) .. | 169 |
| 5.1.5.3. | Borrado (<i>Erase</i>) | 170 |
| 5.1.5.4. | Intersección (<i>Intersect, Clip</i>) | 172 |
| 5.1.5.5. | Diferencia (<i>Difference</i>) | 172 |
| 5.2. | El paquete <i>sf</i> | 174 |
| 5.2.1. | Importación | 176 |
| 5.2.2. | Visualización | 177 |
| 5.2.3. | Proyecciones y coordenadas | 179 |
| 5.2.4. | Acceso, manipulación y selección de variables | 180 |
| 5.2.5. | Funciones de análisis espacial | 182 |
| 5.2.5.1. | Unir tabla (<i>Join</i>) | 183 |
| 5.2.5.2. | Filtrado (<i>Select by attributes</i>) | 184 |
| 5.2.5.3. | Unión (<i>Dissolve</i> y <i>Append</i>) | 185 |
| 5.2.5.4. | Intersección (<i>Intersect, Clip</i>) | 190 |
| 5.2.5.5. | Diferencia (<i>Difference</i>) | 193 |
| 5.2.5.6. | Cálculo de métricas espaciales | 195 |
| 5.2.5.7. | Áreas de influencia (<i>Buffer</i>) | 196 |
| 5.2.5.8. | Cambiar el tipo de geometría | 198 |
| 5.2.6. | Transformación entre formatos: <i>sp</i> , <i>sf</i> , <i>data.frame</i> | 199 |
| 5.3. | Exportación de datos | 200 |
| 6. | DATOS ESPACIO-TEMPORALES II: RÁSTER | 201 |
| 6.1. | Creación de objetos ráster | 201 |
| 6.2. | Importación | 203 |
| 6.3. | Funciones de análisis espacial | 212 |
| 6.3.1. | Álgebra ráster | 212 |
| 6.3.2. | Extraer información con objetos espaciales (<i>Extract to geometry</i>) | 214 |
| 6.3.3. | Modificando el aspecto del ráster | 216 |
| 6.3.4. | Funciones focales | 218 |
| 6.3.5. | Cambio de resolución | 221 |
| 6.3.6. | Reclasificación | 222 |

| | |
|-------------------------------------|-----|
| 6.3.7. Variables topográficas..... | 224 |
| 6.3.8. Otras funciones..... | 227 |
| 6.3.9. Reproyección del ráster..... | 229 |
| 6.4. Exportación de datos | 231 |

PARTE 3.

INTRODUCCIÓN A LA REPRESENTACIÓN CARTOGRÁFICA CON R

| | |
|-----------------------------------------------------------------|-----|
| 7. VISUALIZACIÓN DE DATOS ESPACIALES | 235 |
| 7.1. Representación cartográfica con <i>ggplot2</i> | 235 |
| 7.1.1. Ejemplo: Mapa de accesibilidad media por carretera | 239 |
| 7.1.2. Ejemplo: Mapas de nubosidad en Galicia | 244 |
| 7.1.3. El paquete <i>ggspatial</i> | 254 |
| 7.1.2.1. Imágenes RGB | 254 |
| 7.1.2.2. Escala y norte | 255 |
| 7.1.2.3. Mapas de fondo en teselas..... | 256 |
| 7.1.2.4. Ráster..... | 257 |
| 7.1.4. El paquete <i>ggmap</i> | 258 |
| 7.2. El paquete <i>tmap</i> | 262 |
| 7.2.1. Ejemplo: Presas según tipo de uso | 262 |
| 7.2.2. Ejemplo: Representación del estrés hídrico | 270 |
| 7.2.3. Ráster en <i>tmap</i> | 275 |
| 7.2.4. Exportación con <i>tmap</i> | 277 |
| 7.3. Mapas interactivos..... | 278 |
| 7.3.1. El paquete <i>mapview</i> | 278 |
| 7.3.2. <i>Leaflet</i> | 280 |
| 7.3.3. Exportación de mapas dinámicos | 284 |
| 8. CASOS PRÁCTICOS | 285 |
| 8.1. Visualización del crecimiento urbano | 285 |
| 8.1.1. Importación de datos | 286 |
| 8.1.2. Preparación de los datos..... | 290 |
| 8.1.3. Análisis exploratorio | 291 |
| 8.1.4. Análisis cartográfico | 292 |
| 8.1.5. Mapa dinámico en <i>leaflet</i> | 294 |
| 8.2. Análisis del gasto en I+D. Datos de Eurostat | 295 |
| 8.2.1. Importación de datos | 296 |
| 8.2.2. Análisis exploratorio | 298 |
| 8.2.3. Análisis cartográfico | 300 |
| 8.3. Accesibilidad por carretera en España..... | 302 |

| | |
|---------------------------------------------------------------------|-----|
| 8.3.1. Importación de datos | 303 |
| 8.3.2. Análisis exploratorio | 305 |
| 8.3.3. Análisis cartográfico | 310 |
| 8.3.3.1. Mapa de los datos ráster..... | 310 |
| 8.3.3.2. Mapa de la varianza por provincias | 312 |
| 8.4. Acceso a Open Street Maps desde R | 312 |
| 8.4.1. Acceso a la <i>overpass</i> API desde R y consultas | 313 |
| 8.4.2. Ejemplo: Búsqueda de cines en Madrid | 314 |
| 8.4.3. Ejemplo: Búsqueda de supermercados | 316 |
| 8.5. Cálculo de distancia al mar basada en ráster y vectorial | 318 |
| 8.5.1. La costa de Islandia como ejemplo | 318 |
| 8.5.2. Creación de una red de puntos | 320 |
| 8.5.3. Cálculo de la distancia..... | 321 |
| 8.5.4. Visualización de la distancia calculada | 321 |
| 8.5.5. Exportación de la distancia como ráster | 322 |
| 8.6. Análisis zonal ráster sobre geometrías vectoriales | 324 |
| 8.6.1. Importación y preparación de los datos..... | 324 |
| 8.6.2. Cálculos zonales..... | 325 |
| 8.6.3. Análisis cartográfico | 326 |
| 9. INTEROPERABILIDAD Y EXPORTACIÓN..... | 329 |
| 9.1. SAGA-GIS | 329 |
| 9.1.1. Introducción | 329 |
| 9.1.2. Definir <i>work environment</i> SAGA-GIS..... | 329 |
| 9.1.3. Consultar módulos y funciones de SAGA-GIS..... | 330 |
| 9.1.4. Calcular el <i>Sky View Factor</i> | 332 |
| 9.1.5. Visualización de los resultados | 338 |
| 9.1.6. El paquete <i>lidR</i> | 339 |
| 9.2. <i>rmarkdown</i> | 340 |
| 9.2.1. Primeros pasos | 341 |
| 9.2.2. Opciones globales | 343 |
| 9.2.2.1. Tabla de contenidos (índice) | 343 |
| 9.2.2.2. Resaltado de sintaxis | 343 |
| 9.2.2.3. Opciones de las figuras | 343 |
| 9.2.2.4. Opciones de <i>data.frame</i> | 344 |
| 9.2.2.5. Otras opciones | 344 |
| 9.2.3. Escribir documentos..... | 345 |
| 9.2.3.1. Cabeceras | 345 |
| 9.2.3.2. Formato entre líneas | 345 |

| | |
|-------------------------------------------------------------|-----|
| 9.2.3.3. Listas | 346 |
| 9.2.3.4. Expresiones matemáticas | 347 |
| 9.2.3.5. Referencias en el texto | 347 |
| 9.2.3.6. Código de R..... | 348 |
| 9.2.3.7. Figuras y plots | 349 |
| 9.2.3.8. Tablas | 351 |
| 9.2.4. Creación de documentos desde múltiples archivos..... | 352 |