**INSTITUTO POLITÉNICO NACIONAL**

**UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA Y CIENCIAS SOCIALES Y ADMINISTRATIVAS**

**SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN**

**Maestría en Ciencias en Informática**

**Alumno: Canales Martínez Rubí Verónica**

**Materia: Almacenamiento de Datos y su Administración**

**Profesor: M. en C. Eduardo René Rodríguez Ávila**

**Memoria Técnica**

**Agosto 2016 – Diciembre 2016**

Contenido

[Objetivo 3](#_Toc469524502)

[Descripción 3](#_Toc469524503)

[Análisis 3](#_Toc469524504)

[Desarrollo 3](#_Toc469524505)

[Instalación de R 3](#_Toc469524506)

[Desinstalación de R 9](#_Toc469524507)

[Instalación de *RStudio* 10](#_Toc469524508)

[Desinstalación de *RStudio* 14](#_Toc469524509)

[Instalación de *Git* 15](#_Toc469524510)

[Desinstalación de *Git* 20](#_Toc469524511)

[*GitHub* 21](#_Toc469524512)

[Crear un repositorio en *GitHub* 23](#_Toc469524513)

[*RPubs* 24](#_Toc469524514)

[Conclusiones 29](#_Toc469524515)

[Bibliografía 30](#_Toc469524516)

[Apéndice 30](#_Toc469524517)

# Objetivo

El presente trabajo tiene como objetivo generar una bitácora técnica con la recopilación de herramientas utilizadas por un especialista, administrador o científico para el manejo de datos.

# Descripción

Se busca crear un registro de todo aquello de lo que se haga uso en los aspectos prácticos y aplicativos de la materia de Almacenamiento de Datos y sus Administración; sea una tarea, un ejercicio o por algo que expresamente se deba conocer. Este registro conforma una bitácora técnica de lo que puede considerarse la recopilación de herramientas básicas para todo especialista, administrador o científico enfocado al manejo de datos.

El registro permite conocer como se realizó la selección (si es que hubo), la descarga, instalación y configuración del software empleado, los conceptos básicos y generales de su uso e instrucciones de desinstalación.

Su organización se lleva a cabo "por tiempo" honrando el concepto de bitácora por orden de instalación o uso en el tiempo.

# Análisis

La importancia de generar una memoria técnica se basa en la comprensión de la descripción del curso de Almacenamiento de Datos y su Administración impartido en la Unidad Profesional Interdisciplinaria de Ingeniería y Ciencias Sociales y Administrativas, dicho curso “presenta los elementos de sustento formales alrededor de los principios teóricos y prácticos en las acciones y para las aplicaciones destinadas a resguardar y recuperar elementos de información que son el insumo para y el resultado de la operación de sistemas de cómputo e informáticos. Una serie de ejercicios alrededor de los temas presentados permiten apreciar el aspecto práctico de la materia1”.

El objetivo del curso es que al término del mismo el estudiante contará con habilidades y conocimientos específicos para manejar conjuntos de datos de diversos tamaños bajo un panorama que contempla su adquisición, adecuada preparación y óptimo procesamiento con el fin de generar e identificar los resultados adecuados que propicien el descubrimiento de nuevos datos e información. Se buscará además, que el estudiante desarrolle un criterio para cultivar mayores habilidades que le permitan, inclusive, la administración de grandes volúmenes de datos1.

# Desarrollo

## Instalación de R

A lo largo del curso de Almacenamiento de Datos y su Administración se hará uso del lenguaje de programación R que proporciona un enfoque al desarrollo estadístico, es un software libre que utiliza modelos estadísticos, de visualización y bioinformáticos (es la aplicación de tecnologías computacionales a la gestión y análisis de [datos](https://es.wikipedia.org/wiki/Dato) [biológicos](https://es.wikipedia.org/wiki/Biolog%C3%ADa)).

A continuación se describirá la instalación de este lenguaje de programación:

1. Bajar de la siguiente página *web* R 3.3.2 for *Windows* (62 megabytes, 32/64 bit)

<https://cran.r-project.org/bin/windows/base/>

En dicha página también se puede encontrar información sobre la instalación y las nuevas características de esta versión.

1. Dar click en el link: *Download R 3.3.2 for Windows*

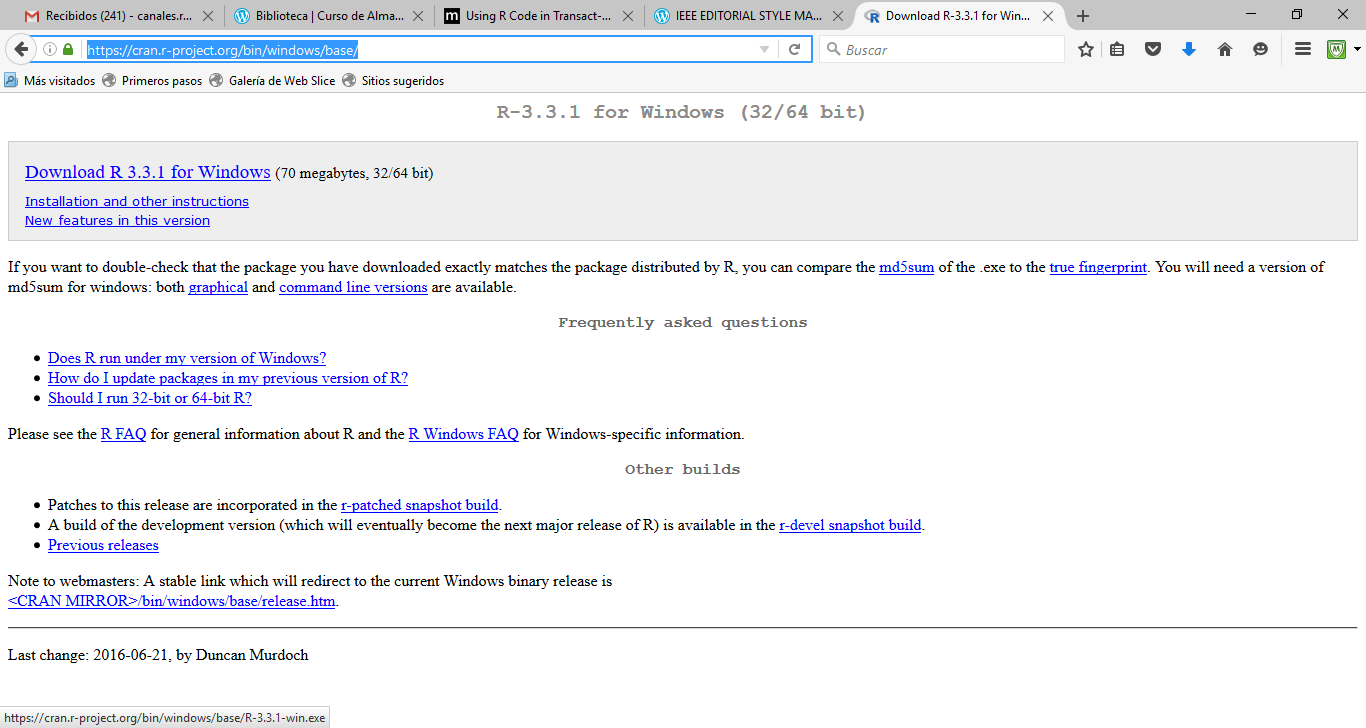
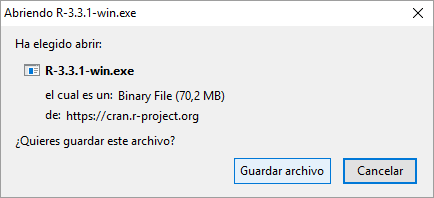
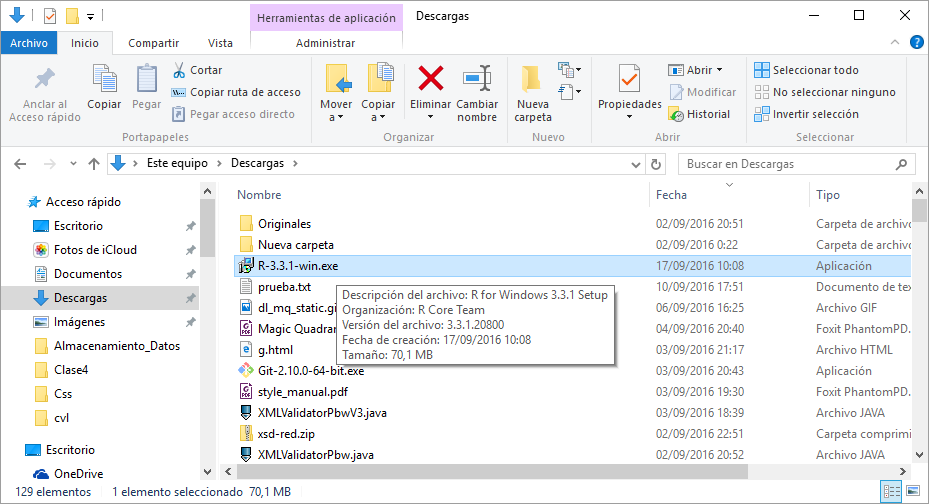


Figura 1. Página de donde se puede obtener R oficialmente.

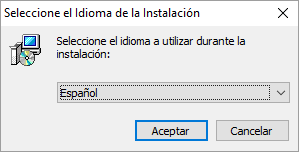
1. Enseguida aparecerá la pantalla de descarga del *software*, preguntándonos que si queremos guardar el archivo, a lo que debemos responder “Guardar archivo”:



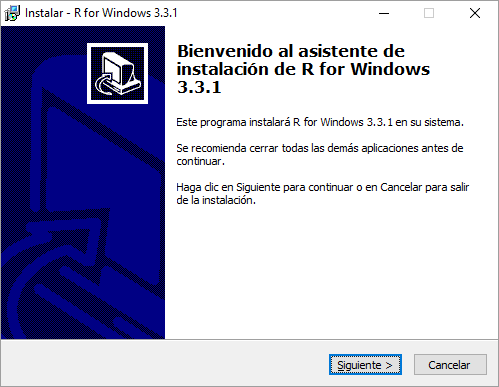
1. Una vez que hayamos bajado el archivo, buscarlo en la carpeta de descargas y dar doble clic para ejecutarlo:



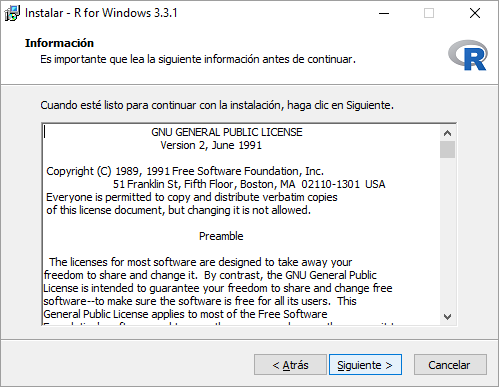
1. La siguiente pantalla que nos aparecerá es la selección del idioma sobre el que queremos instalar R, elegimos español, si así lo deseamos.



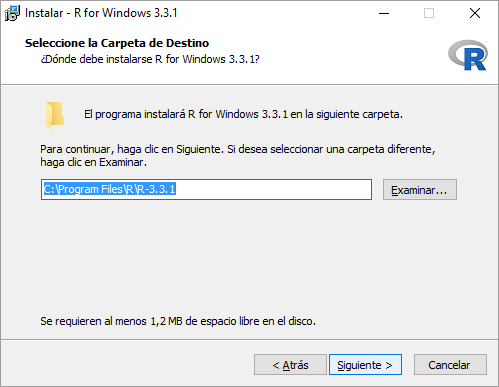
1. Enseguida se mostrará una pantalla de asistente de instalación de R, damos siguiente para continuar:



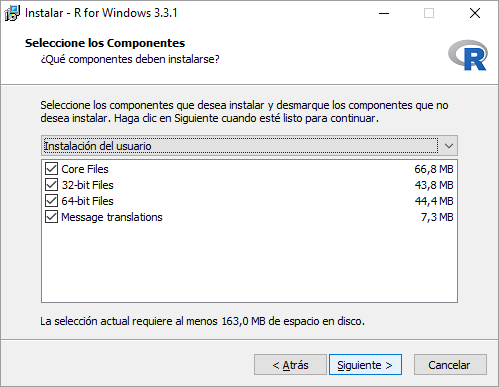
1. La siguiente pantalla muestra información de la licencia y otros datos importantes del software de R:



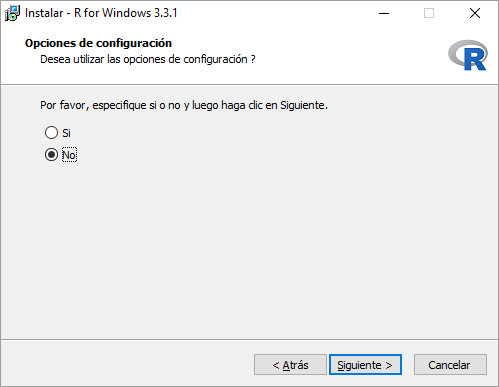
1. La siguiente pantalla nos pregunta sobre la ruta donde se instalará el *software* R, se recomienda no cambiar la ruta de instalación por default y dar siguiente para continuar:

****

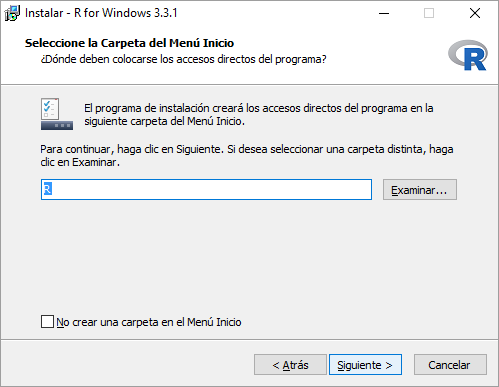
1. La siguiente pantalla nos pregunta sobre los componentes que queremos instalar de R, dar siguiente para continuar:



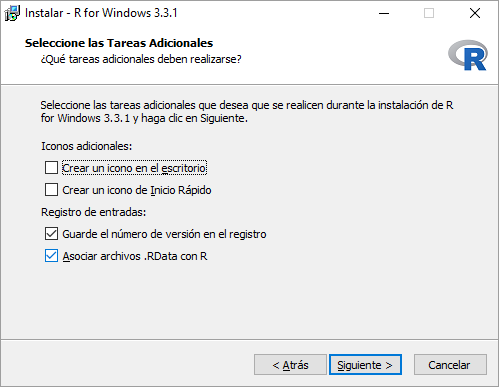
1. La siguiente pantalla nos pregunta si deseamos utilizar las opciones de configuración?, damos no y siguiente para continuar:

****

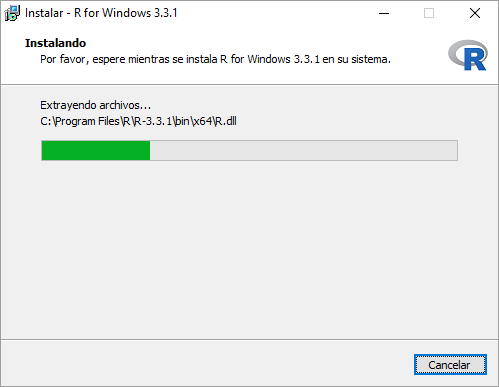
1. La siguiente pantalla nos pregunta ¿dónde deben colocarse los accesos directos del programa? Dejamos la opción por default y darnos siguiente para continuar:

****

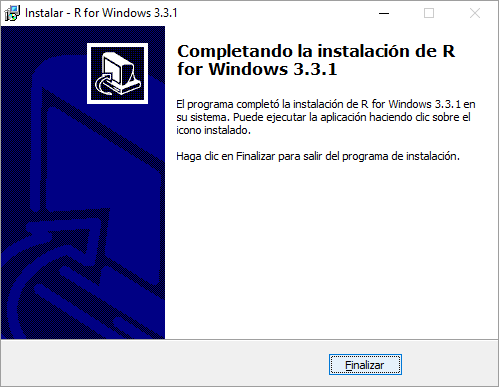
1. La pantalla siguiente nos permite elegir entre las opciones de crear iconos en el escritorio y de inicio rápido, así como la asociación de los archivos de tipo R con la extensión .*RData*. Seleccionamos nuestra personalización deseada y damos siguiente para continuar:

****

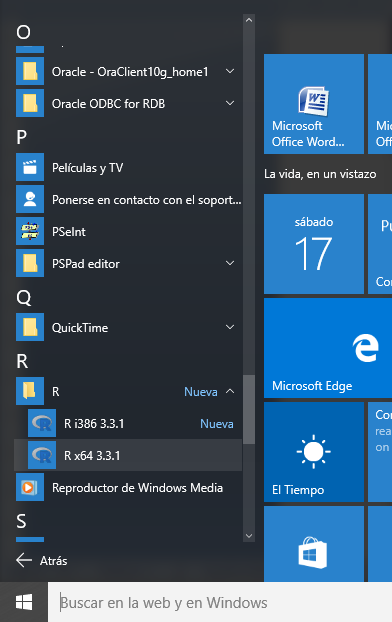
1. A continuación se mostrará una pantalla que nos mostrará el avance de la instalación:

****

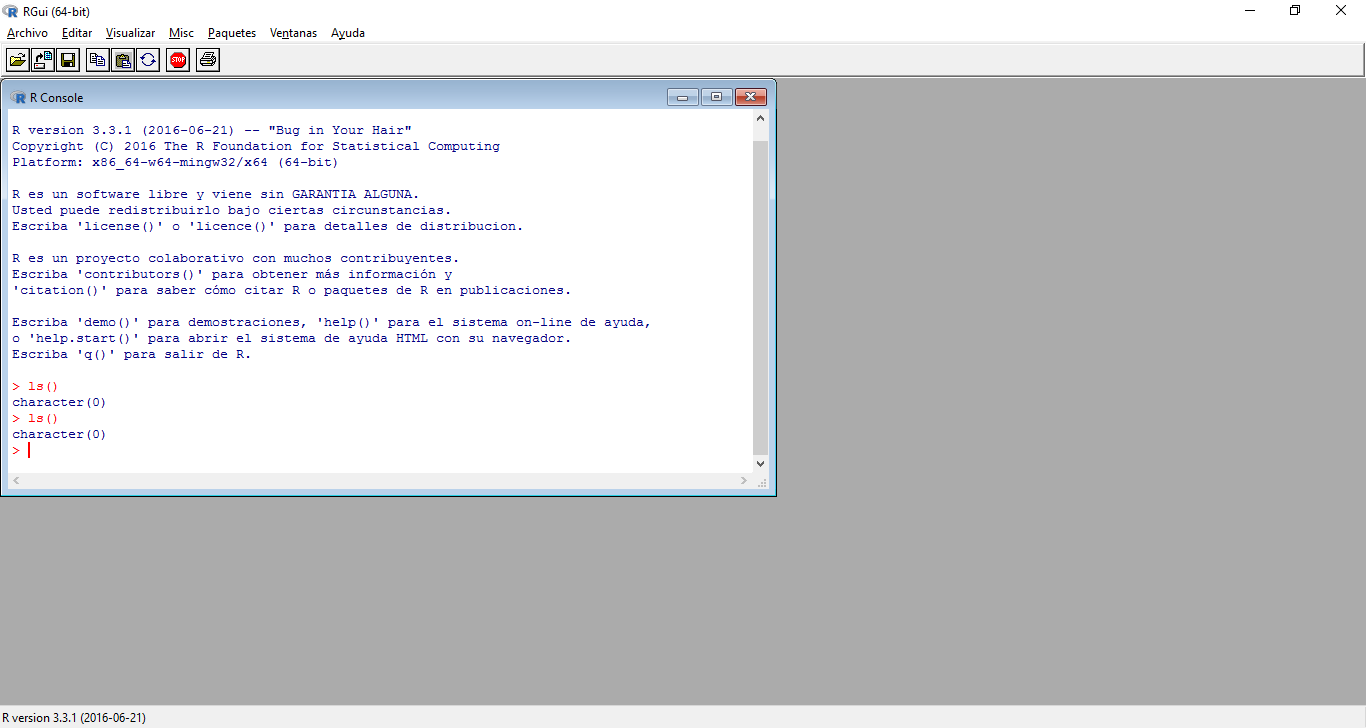
1. Finalmente nos mostrará la pantalla de finalización de la instalación:

****

1. Se recomienda revisar la instalación ejecutando el programa para cerciorarnos que la instalación fue correcta:

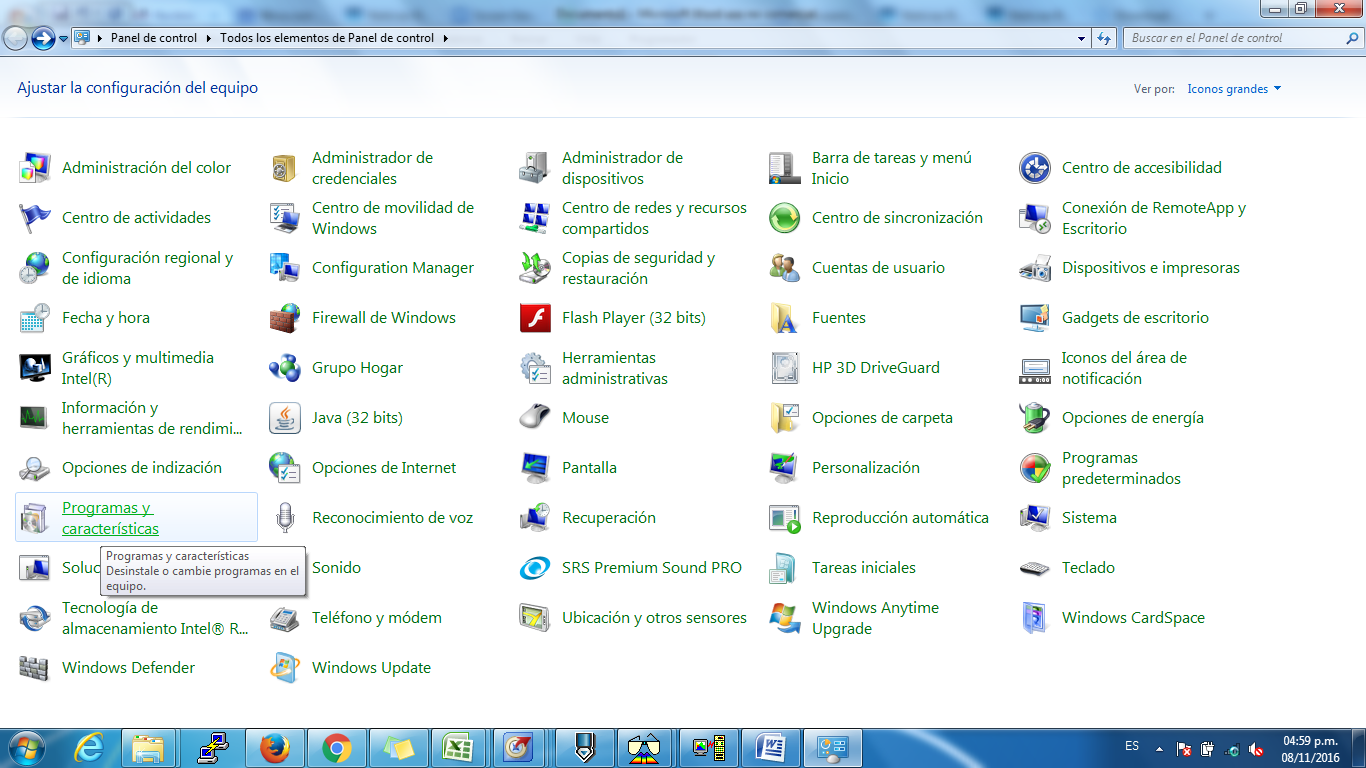
****

1. Una vez que el programa se habrá, mostrará la siguiente pantalla:

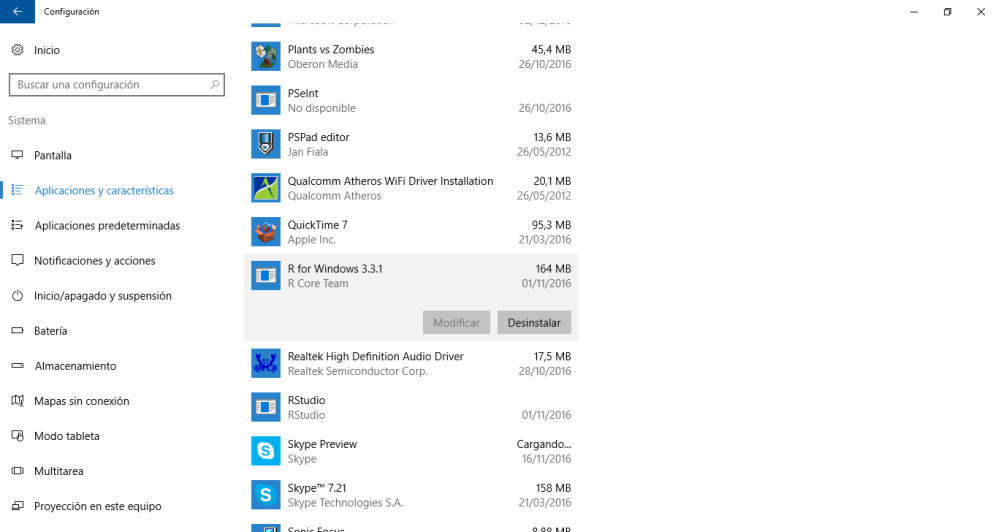
****

## Desinstalación de R

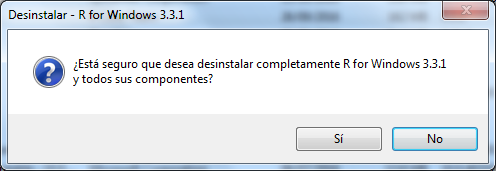
Ir a Inicio 🡪 Panel de control 🡪 Programas y características:



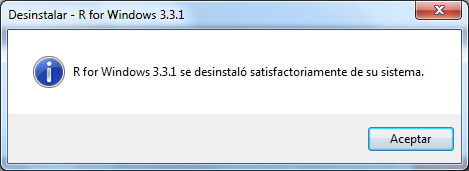
Seleccionar *R for Windows 3.3.1* y dar clic en Desinstalar:



Nos preguntará nuevamente para confirmar si deseamos desinstalar, elegir Sí:



Al terminar la desinstalación enviará el mensaje de confirmación de finalización:



## Instalación de *RStudio*

Ahora bien, R actualmente cuenta con un software GUI (interfaz gráfica de usuario) *R Studio* que nos permite hacer más fácil el uso de R, incluye un editor de código, herramientas de visualización y *debuggeo* de código. Dicho software podemos obtenerlo de la página:

<https://www.rstudio.com/products/rstudio/>

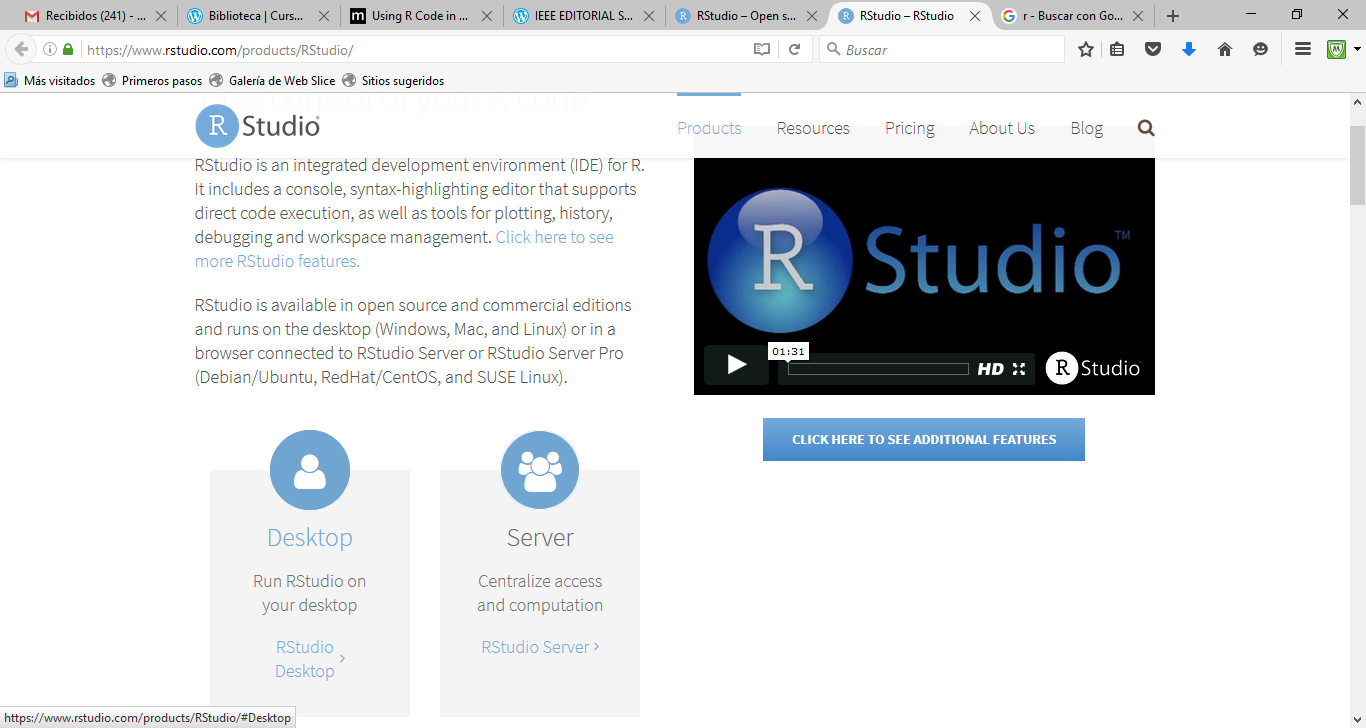
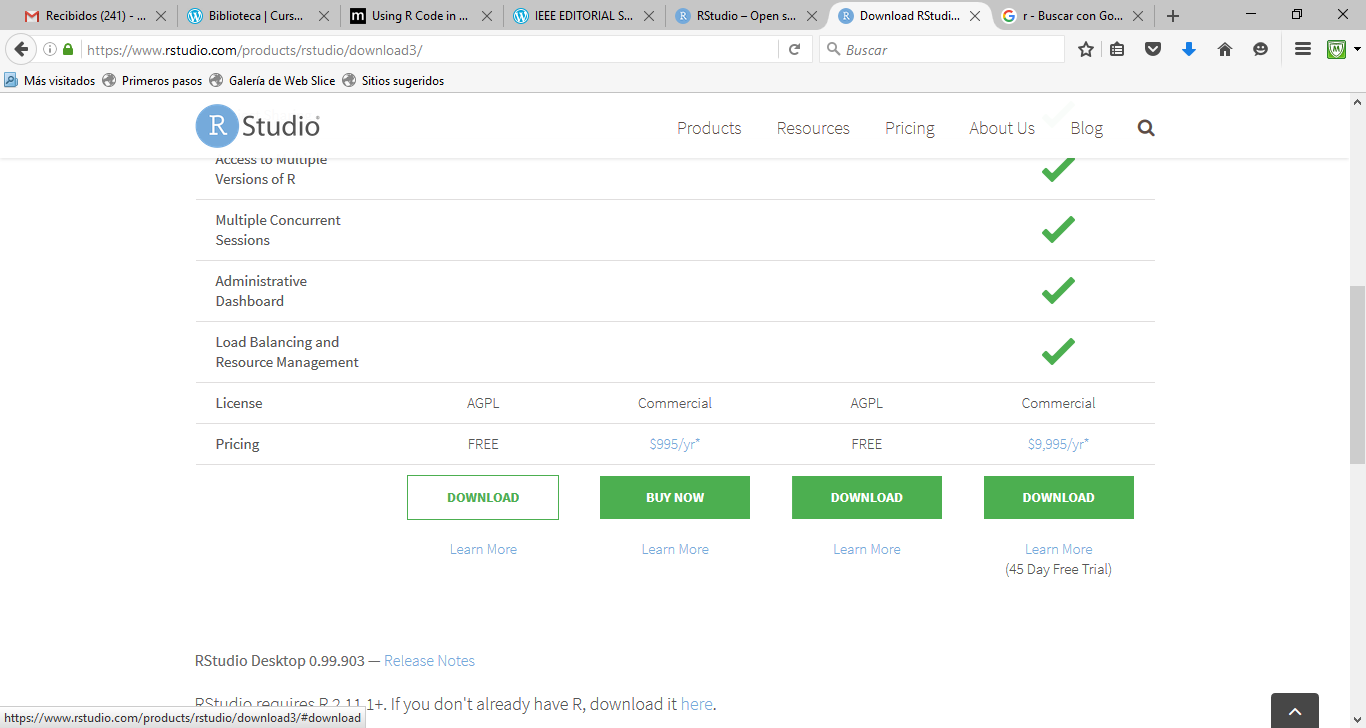
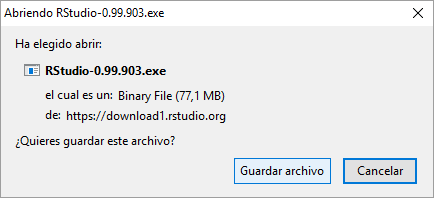
****

Figura 2. Página de donde se puede obtener R Studio oficialmente.

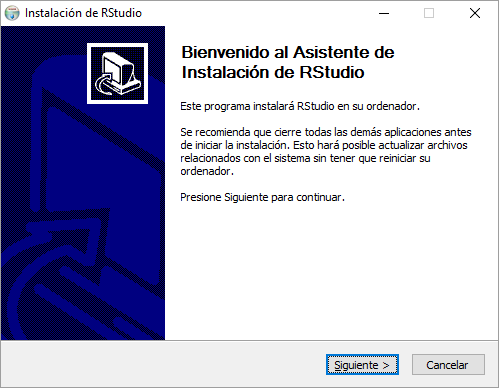
1. Elegimos la opción *Desktop* y seleccionamos la versión con fines educativos (*free*):

****

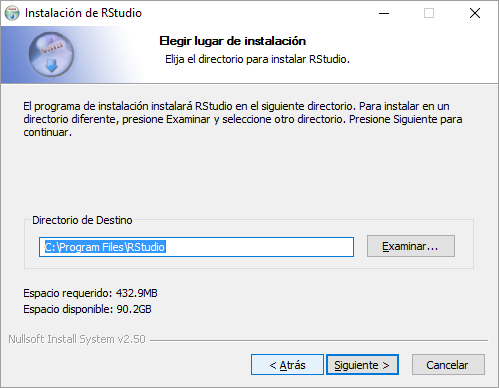
1. Nos preguntará si queremos guardar el archivo, elegimos Guardar archivo:



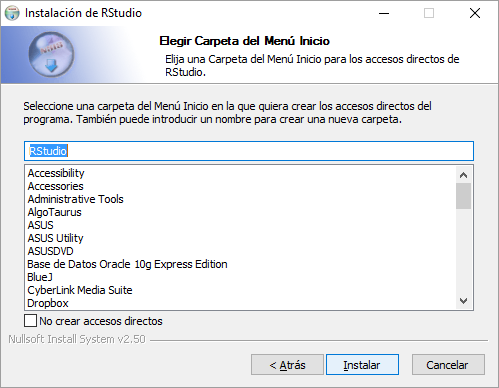
1. Ejecutamos el archivo *RStudio*-0.99.903.exe. Y en seguida nos enviará la siguiente pantalla de inicio de la instalación y damos siguiente para continuar:

****

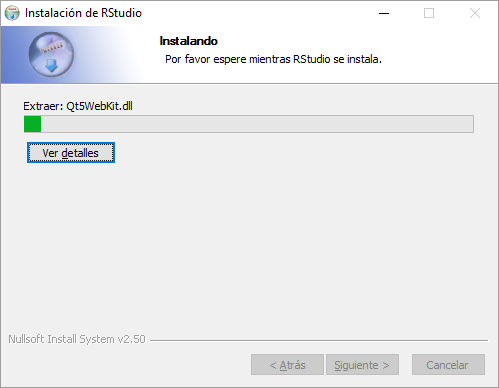
1. Ahora nos preguntara en que directorio deseamos instalar, se recomienda dejar la de default y dar siguiente para continuar:

****

1. A continuación nos indicará el nombre del acceso directo para *RStudio*, damos Instalar para continuar:

****

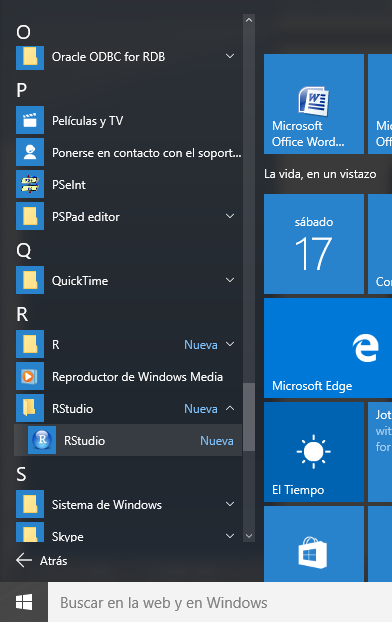
1. A continuación nos mostrará una pantalla con el avance de la instalación:

****

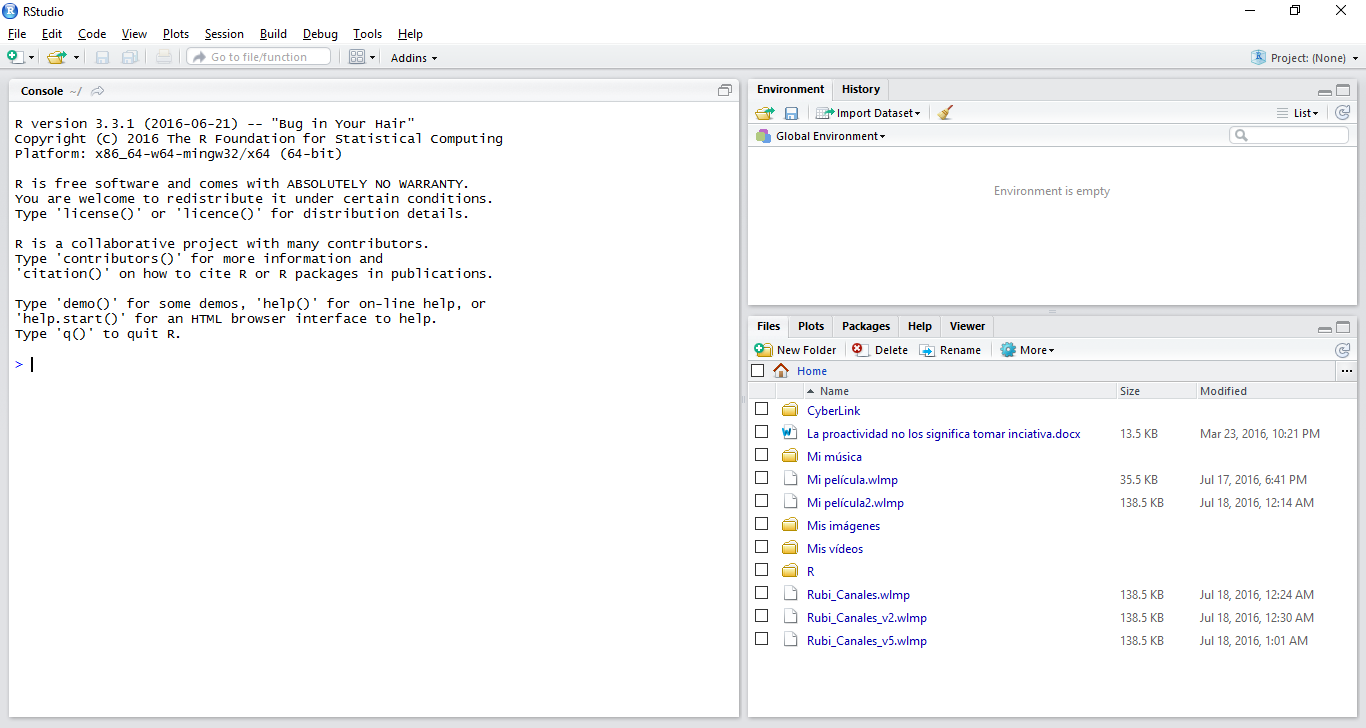
1. Finalmente, nos enviará la siguiente pantalla de finalización de la instalación de *RStudio*:

****

1. Para garantizar la correcta instalación de *RStudio*, lo ejecutamos:

****

1. Una vez ejecutado nos abrirá la siguiente pantalla que nos permitirá comenzar a interactuar con *RStudio*:

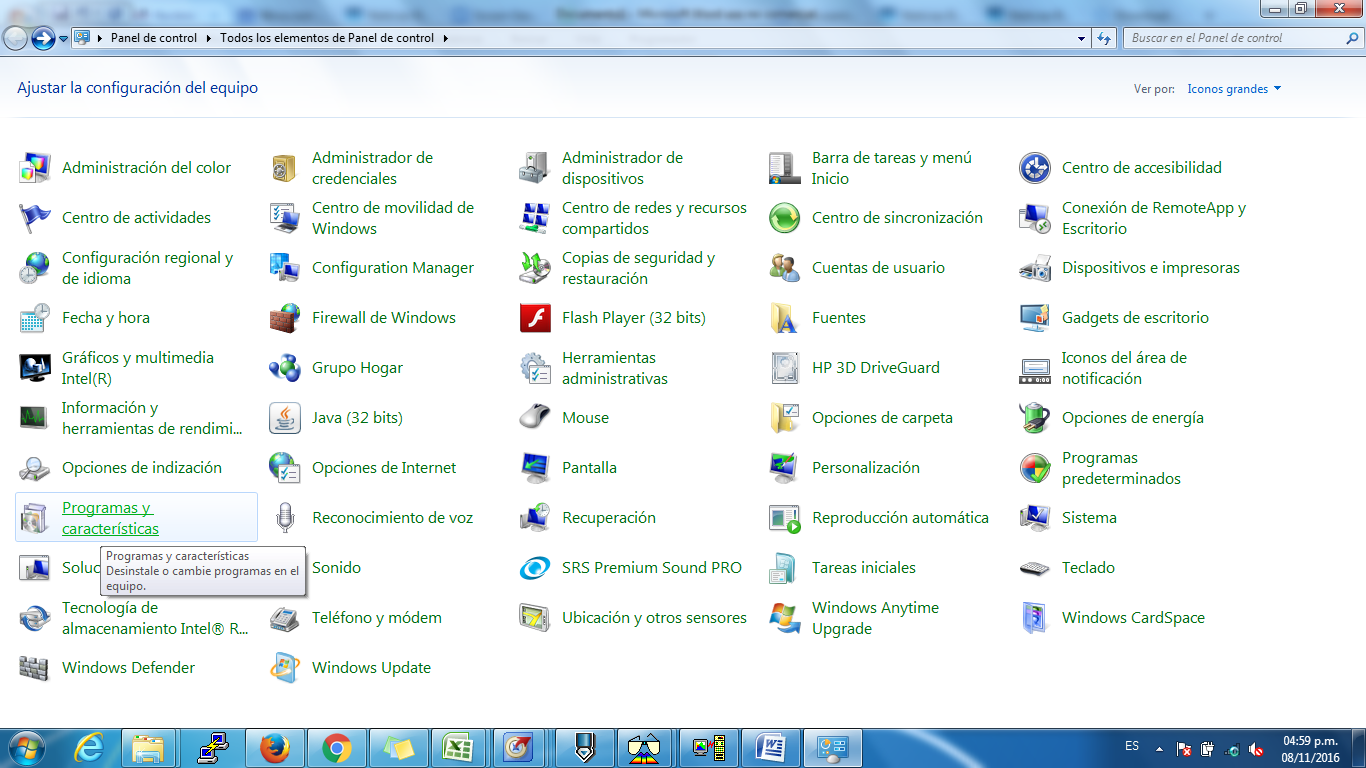
****

Como se menciono al inicio de las instrucciones de instalación RStudio es mucho más amigable por ser visual y permitir a través de sus pantallas interactuar con él; así las pantallas más representativas son:

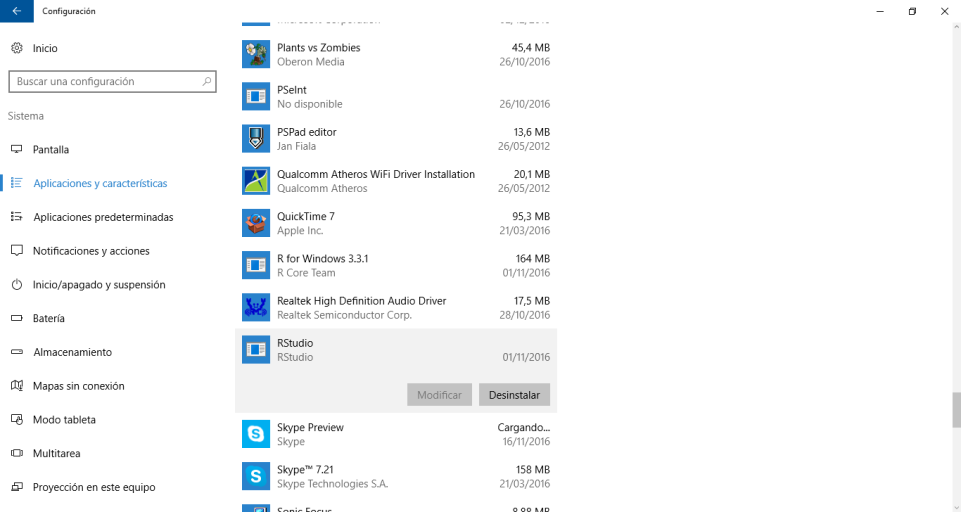
* *Console*.- en ella podemos escribir y ejecutar el código R.
* *Environment*.- nos mostrará las variables y/o funciones que vayamos declarando.
* *History*.- nos muestra el código que se va ejecutando como un historial de las instrucciones ejecutadas.
* *Files*.- nos muestra el directorio donde estamos trabajando.
* *Packages*.- las bibliotecas que podemos instalar.
* *Help*.- la ayuda de los comandos que vayamos solicitando en la consola.
* *Viewer*.- es el visualizador de gráficas.

## Desinstalación de *RStudio*

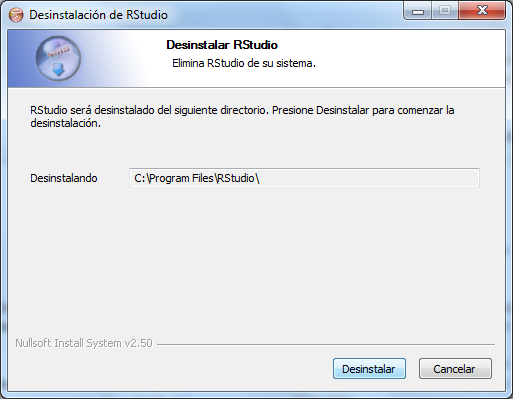
Ir a Inicio 🡪 Panel de control 🡪 Programas y características:



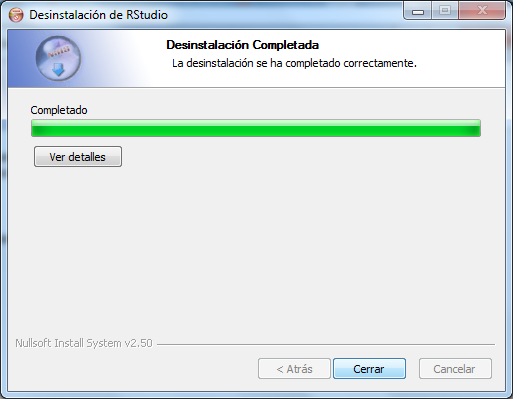
Seleccionar *RStudio* y dar clic en Desinstalar:



Comenzará el proceso de desinstalación, seleccionar Desinstalar:

****

Mostrará la pantalla del proceso de avance:

****

## Instalación de *Git*

*Git* es un software de control de versiones pensando en la eficiencia y la confiabilidad del mantenimiento de versiones de aplicaciones cuando éstas tienen un gran número de archivos de código fuente. *Git* le da a cada programador una copia local del historial del desarrollo entero, y los cambios se propagan entre los repositorios locales.

Esta herramienta gestiona los cambios de un conjunto de archivos. Cada conjunto de cambios genera una nueva versión de los archivos. El controlador de versiones permite, además de la gestión de cambios, recuperar una versión vieja de los archivos o un archivo, así como resolver conflictos entre versiones.

Aunque su principal uso es para agilizar la colaboración en el desarrollo de proyectos de software, también puede utilizarse para otros fines (como notas) e inclusive para trabajar en proyectos.

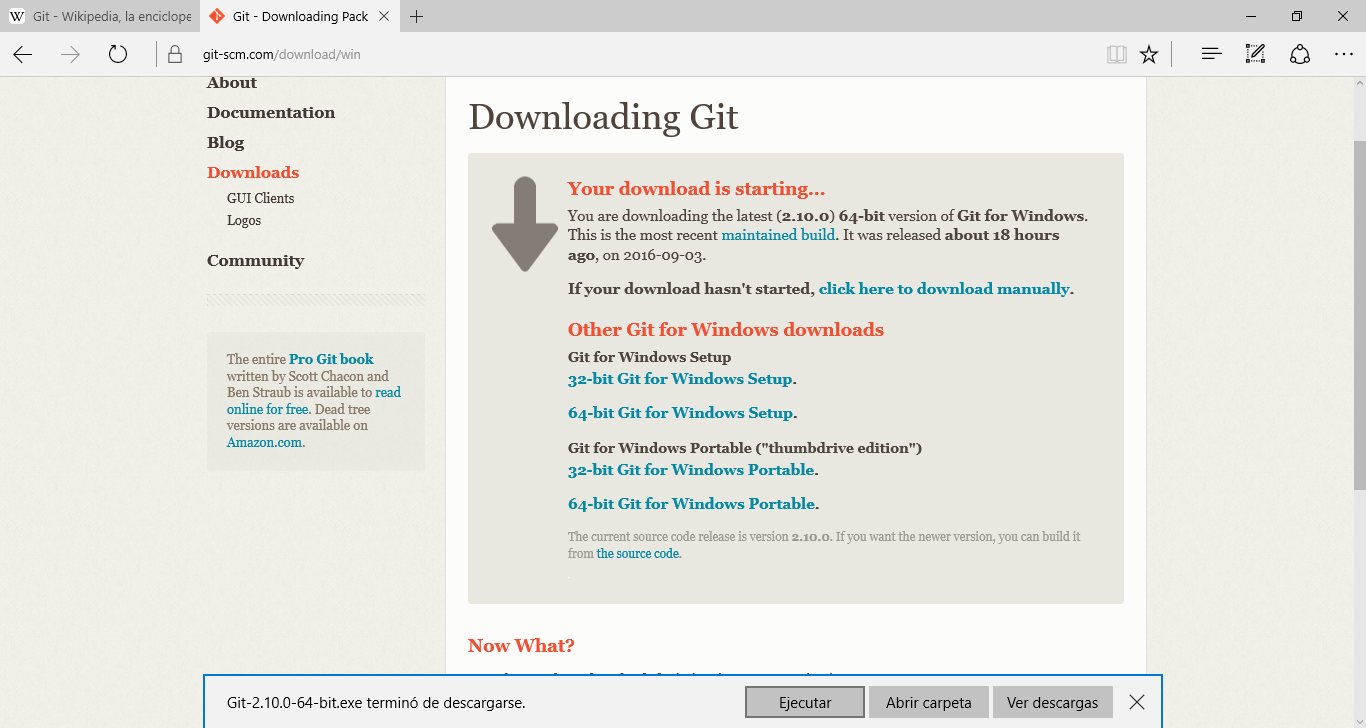
*Git* se puede obtener de la siguiente página:

<https://git-scm.com/downloads>

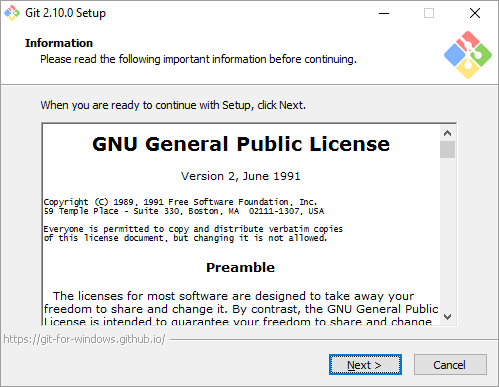


Figura 3. Muestra la página de donde se puede bajar *Git.*

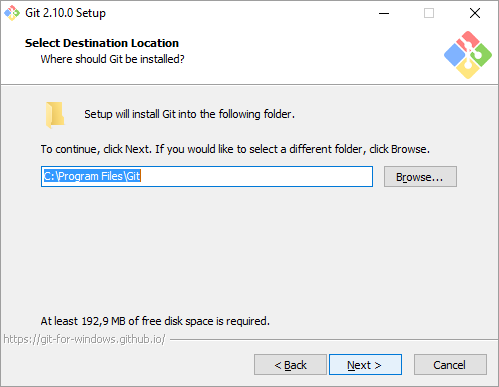
1. La instalación comienza bajando la versión a instalar según el sistema operativo que se tenga, en este caso Windows:



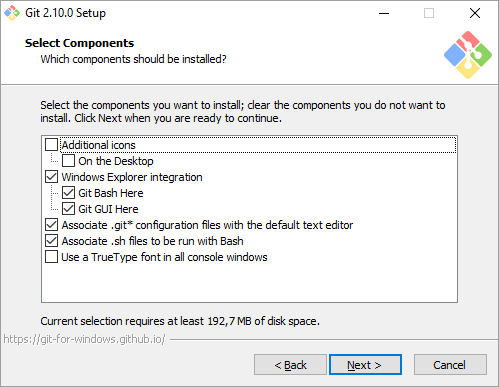
1. Ejecutar el archivo Git-2.10.0-64-bit.exe y nos aparecerá la siguiente pantalla con el inicio de la instalación. Se mostrará datos de la licencia y del *software* en sí, dar *Next* para continuar:



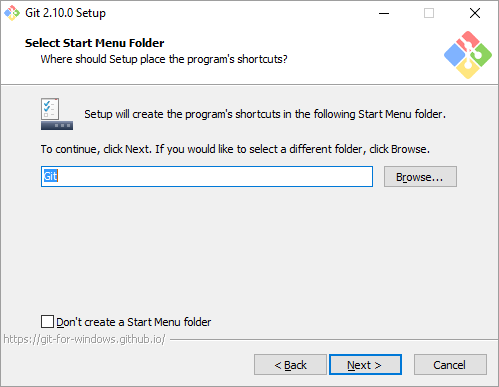
1. Después aparecerá una pantalla que nos indica donde se instalará el software, se recomienda no cambiar la ruta y después dar *Next* para continuar:



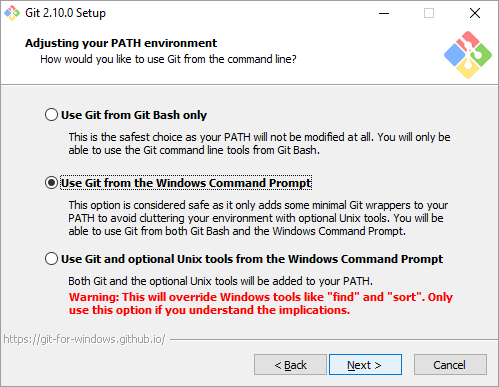
1. Aparecerá la siguiente pantalla, donde nos pregunta que componentes queremos instalar con *Git*, seleccionar los deseados y dar *Next* para continuar:



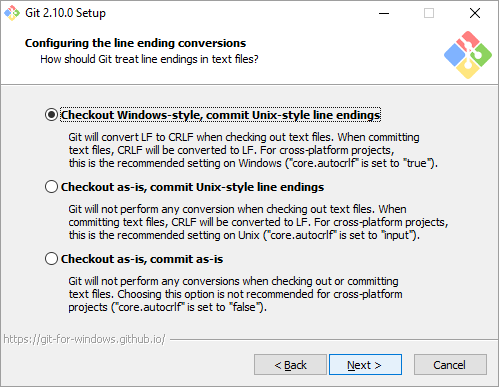
1. Después aparecerá una pantalla indicándonos en que folder se instalará el software, se recomienda no cambiarlo y dar *Next* para continuar:



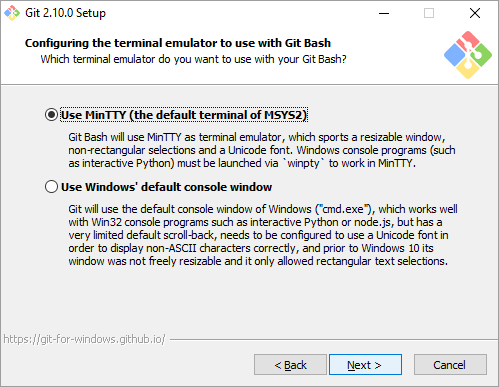
1. Enseguida aparecerá una pantalla preguntando si nos gustaría usar *Git* desde la línea de comandos, seleccionar esa opción y dar *Next* para continuar:



1. Enseguida nos preguntará sobre la configuración de *Git*, seleccionar la opción predefinida y dar *Next* para continuar:



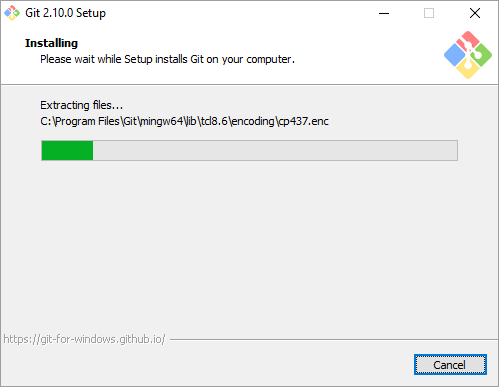
1. Nuevamente nos hará otra pregunta sobre la configuración de *Git Bash*, dejar la *default* y dar *Next* para continuar:



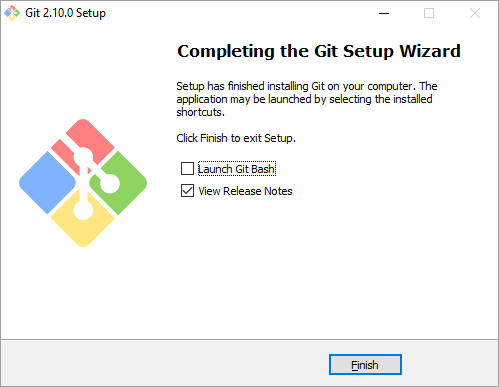
Ahora habrá que elegir las características de *Git* que se quiera habilitar, dejar las de *default* y dar *Next* para continuar:



1. Comenzará la instalación mostrándonos una barra de avance:

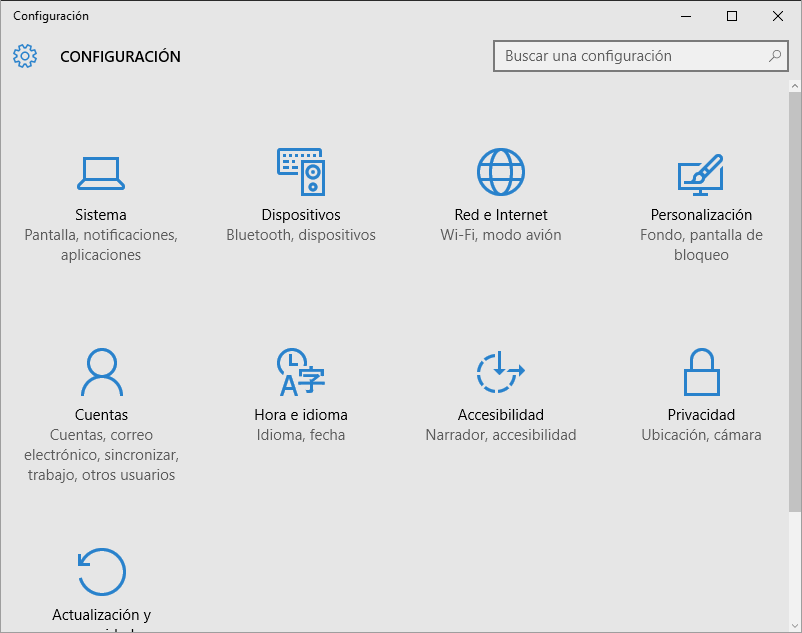


1. Finalmente nos mostrará la pantalla de finalización de la instalación de *Git,* dar *Finish* para terminar:

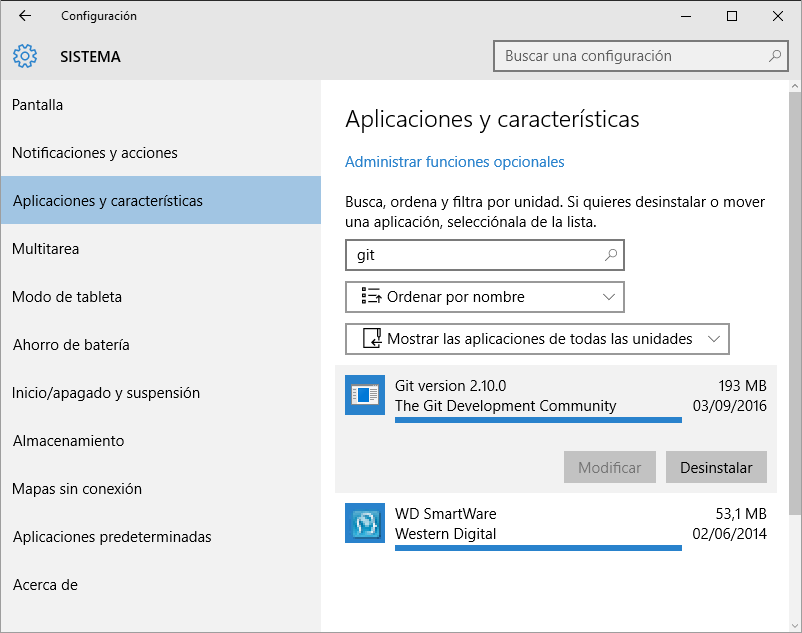


## Desinstalación de *Git*

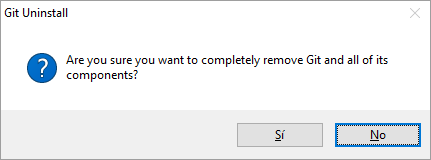
1. En la barra de Inicio, ir a Configuración y seleccionar Sistema:



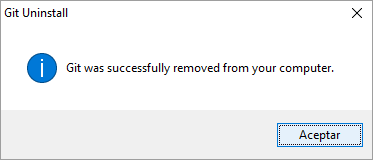
1. Después elegir Aplicaciones y características y seleccionar *Git* :



1. Seleccionar Desinstalar y nos enviará un mensaje de confirmación sobre la desinstalación:



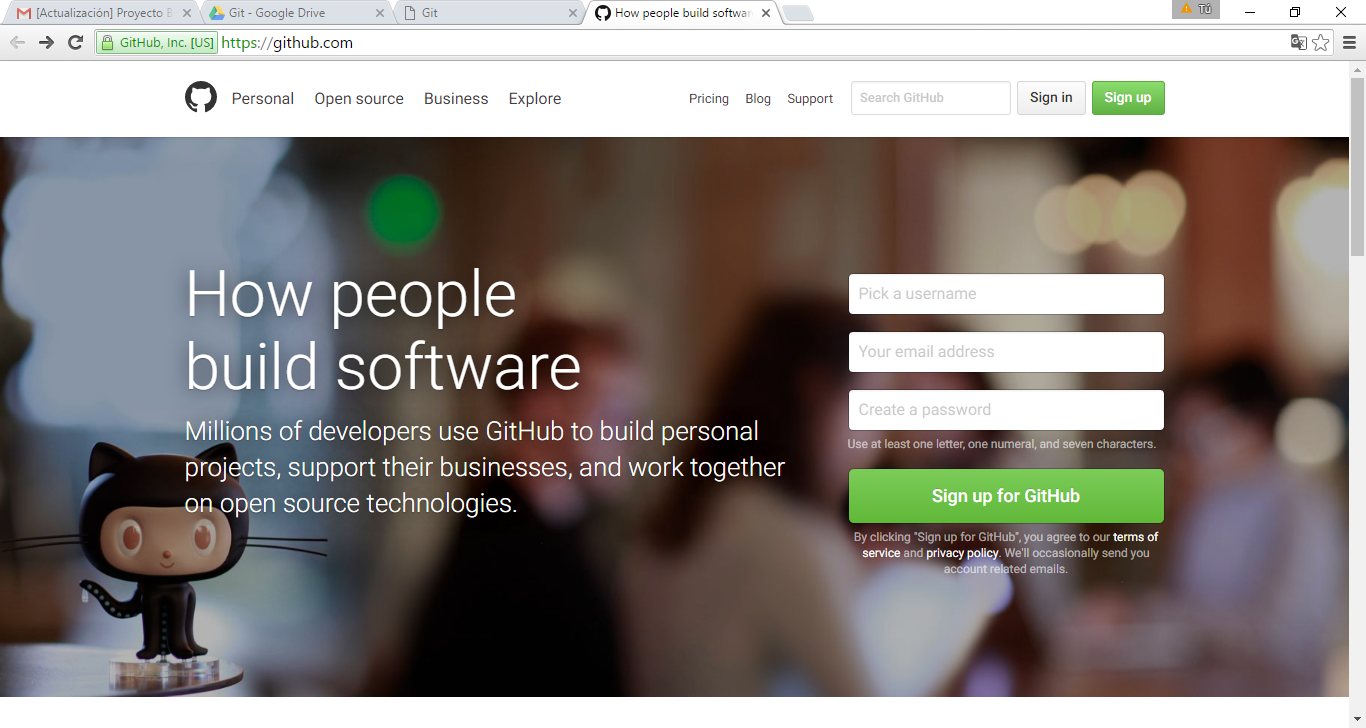
1. Una vez finalizada la instalación nos enviará un mensaje de que ha terminado con la desinstalación:

****

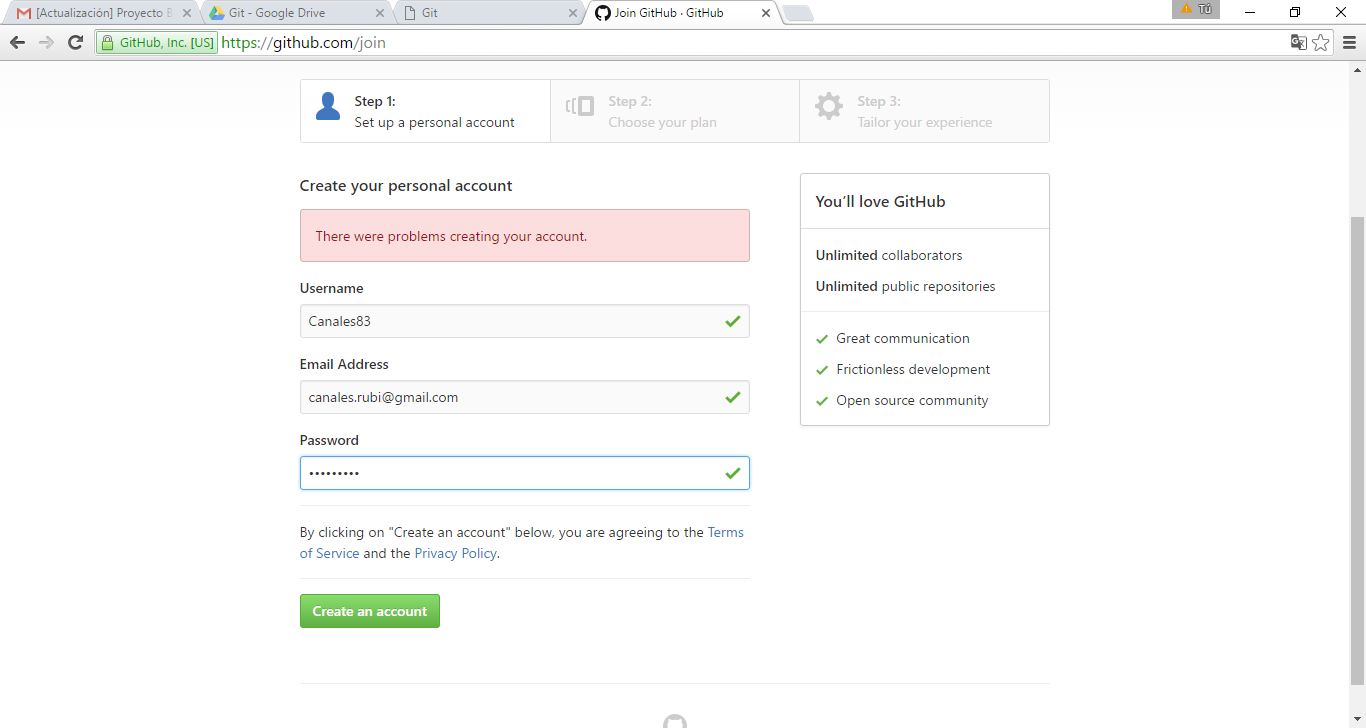
## *GitHub*

Ahora vamos a crear una cuenta en *GitHub* accesando a la página:

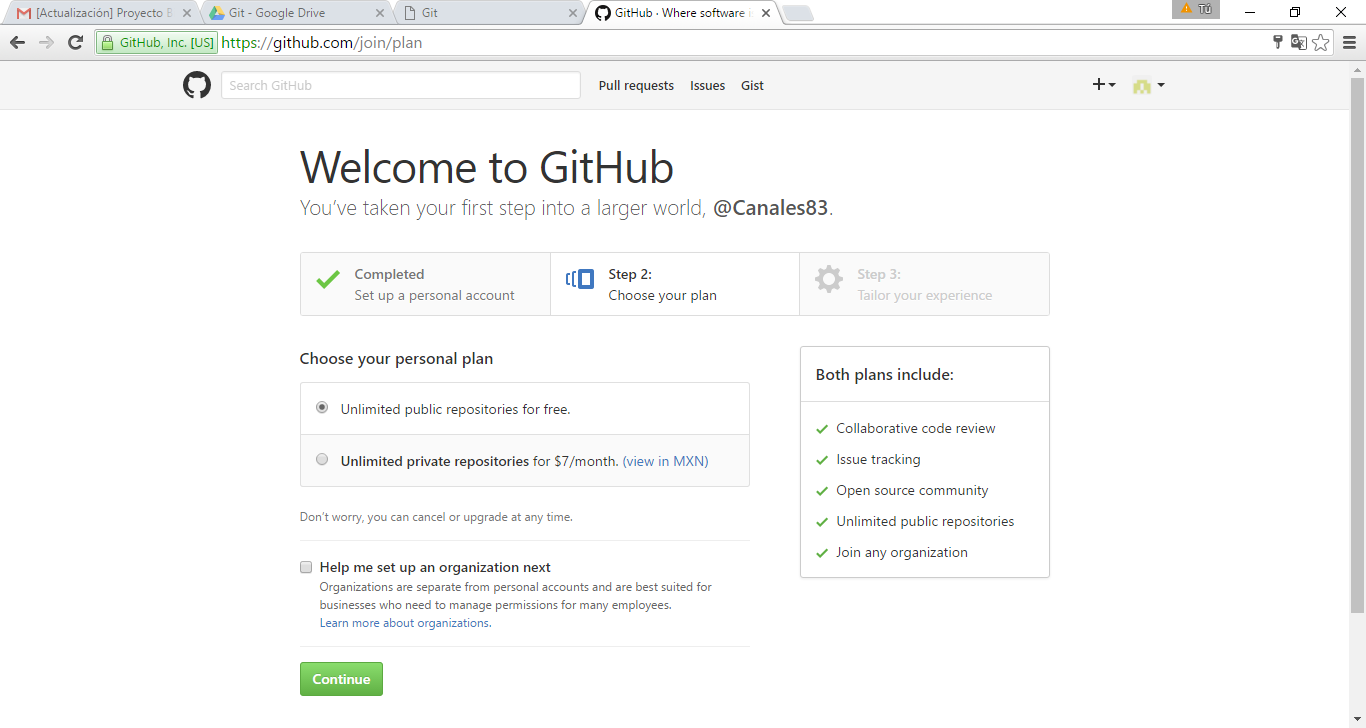
<https://github.com/>



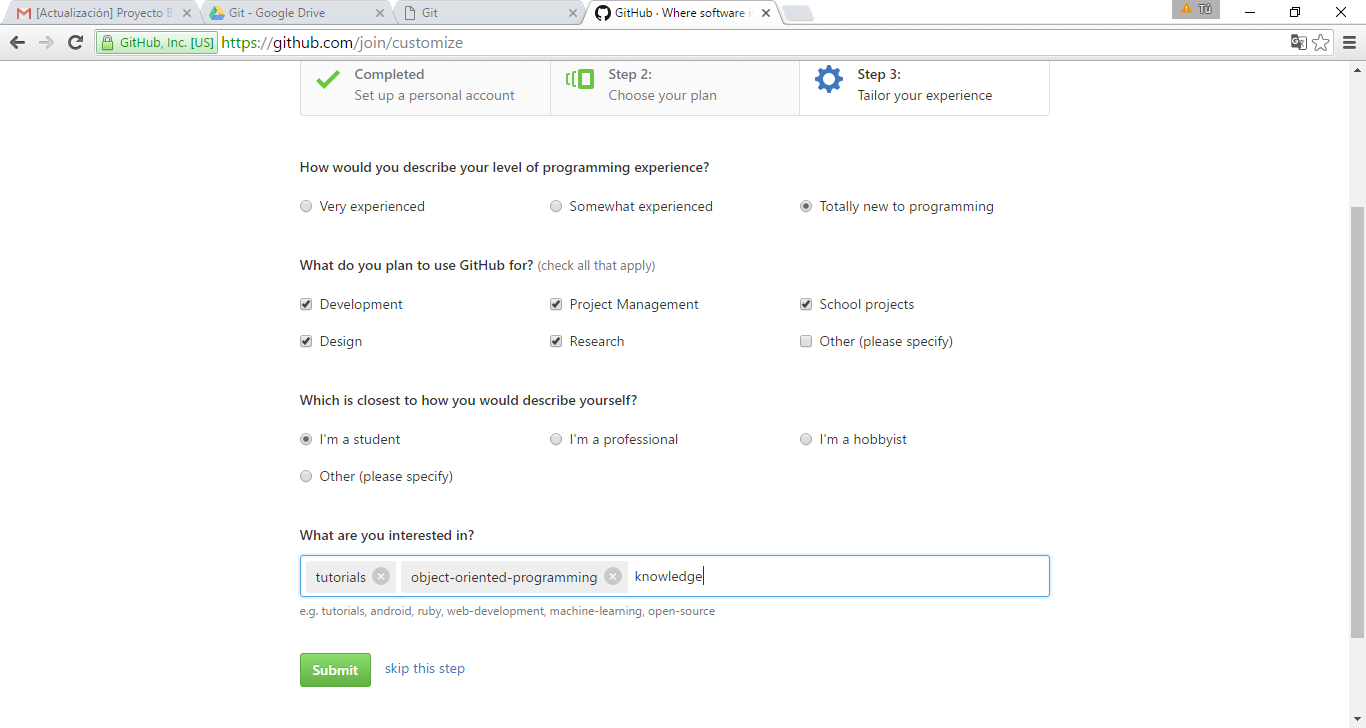
1. Elegir ***Sign up for GitHub*** para crear una cuenta, nos pedirá un nombre de usuario, un email a donde asociarlo y una contraseña y seleccionar ***Create an account***:



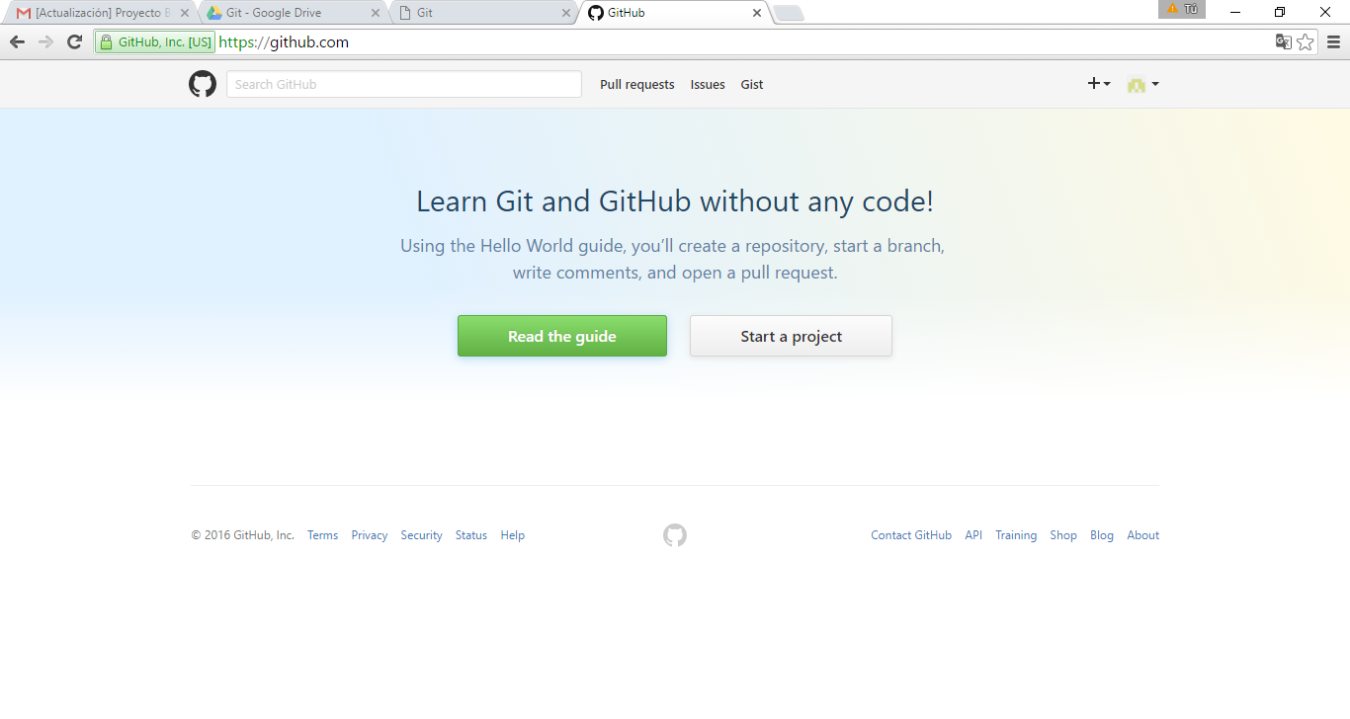
1. El paso 2 de la creación de la cuenta consiste en elegir la versión libre o la por pagar. La diferencia radica en que la de paga permitirá principalmente crear repositorios privados, mientras que la libre cualquier usuario puede accesar a ella:



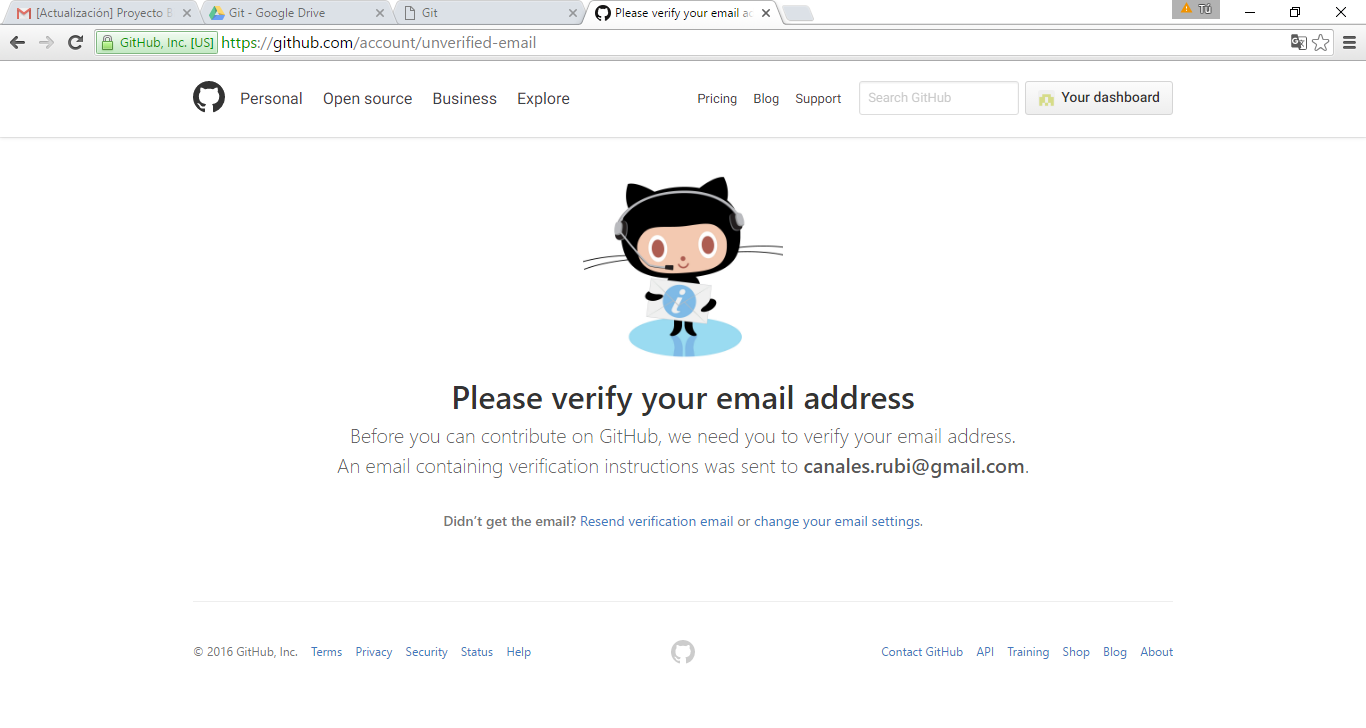
1. El paso 3 consiste en complementar información personal como el nivel de experiencia en programación, el uso que le daremos a *GitHub*, tu ocupación y tus intereses; seleccionar *Submit* para continuar:



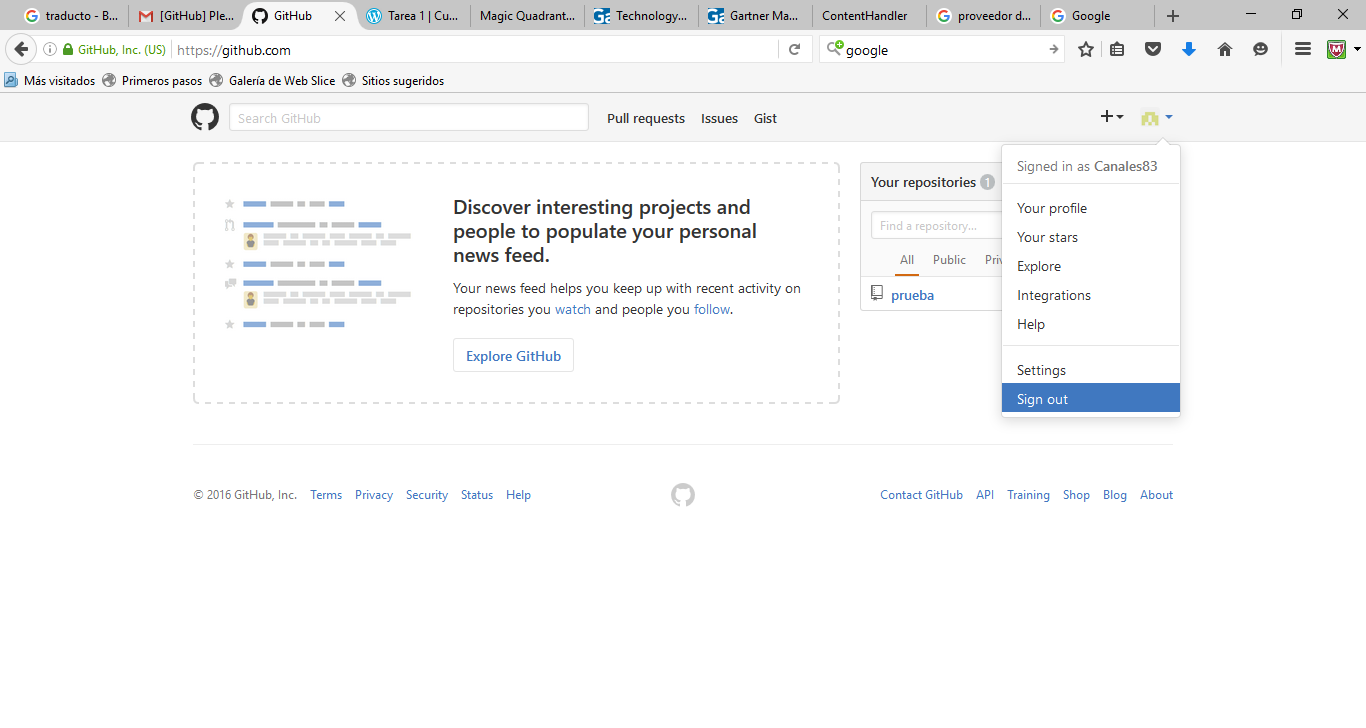
1. Una vez concluida esta etapa nos mandará a una página para que comencemos a trabajar con *GitHub*, o bien, leer la guía de uso:



1. Antes de comenzar a utilizar GitHub, habrá que verificar nuestra cuenta, accesando al correo que dimos de alta y con ello confirmar nuestra inscripción a *GitHub*:



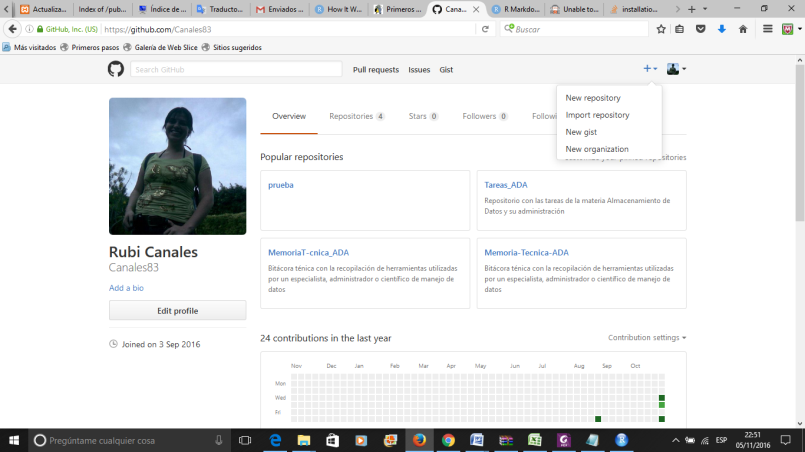
1. Para salir de nuestra cuenta en *GitHub* basta con dar clic en la flecha superior derecha y desplegará un menú, seleccionar *Sign out*:



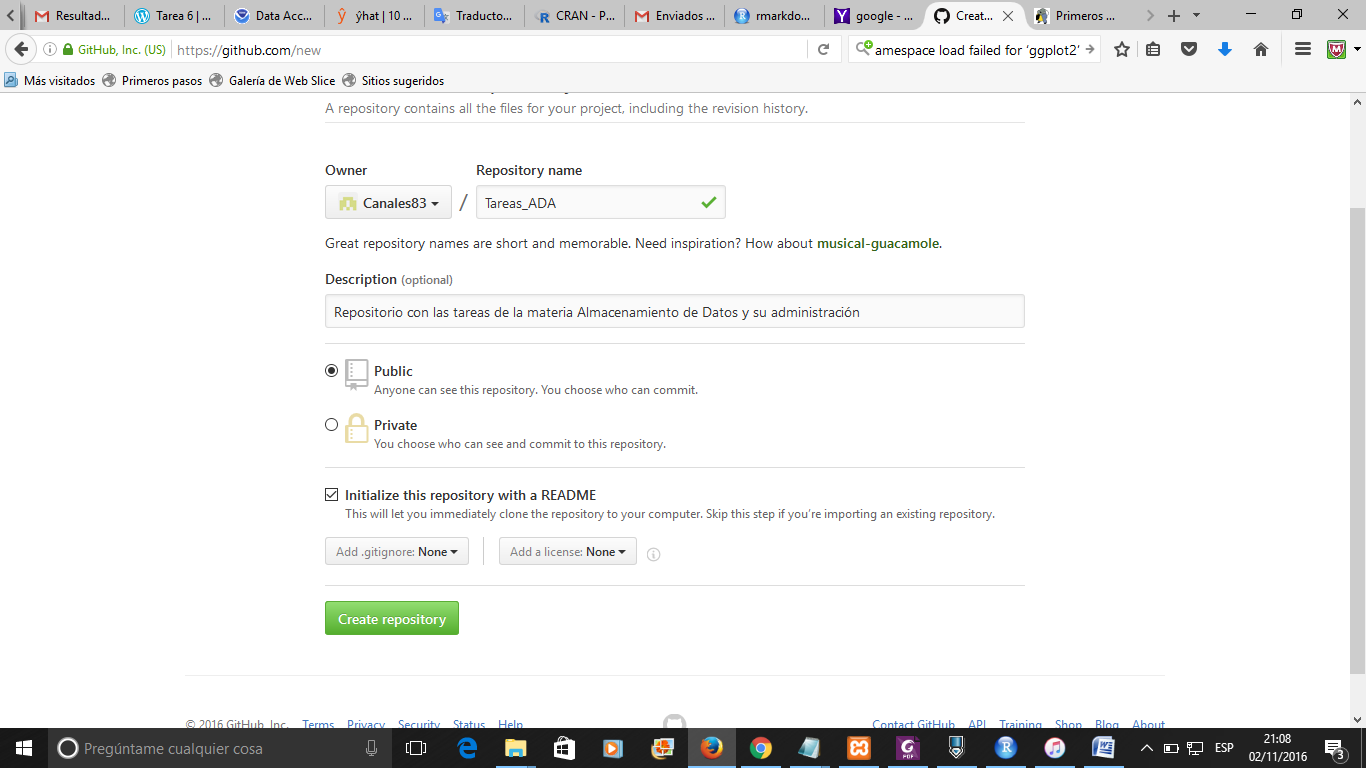
## Crear un repositorio en *GitHub*

**Un Repositorio es un** directorio, carpeta o lugar donde viven los proyectos. Puede estar ubicado en tu computadora, en GitHub u en otro host. En él se almacén códigos, archivos de texto, imágenes que Git guardará y gestionará para todas las versiones que se vayan haciendo de ese archivo.

En la página principal del *GitHub* seleccionar en el icono de **+** posicionado en la parte superior derecha y seleccionar *New repository*:



Nos solicitará datos como: nombre del repositorio, descripción, será público o privado (recordando que si elegimos privado deberemos pagar una cuota para poder hacerlo privado), y podemos seleccionar Inicializar el repositorio con un archivo README.md.



Una vez creado ya podemos hacer uso del repositorio para poder subir nuestros archivos en él.

## *RPubs*

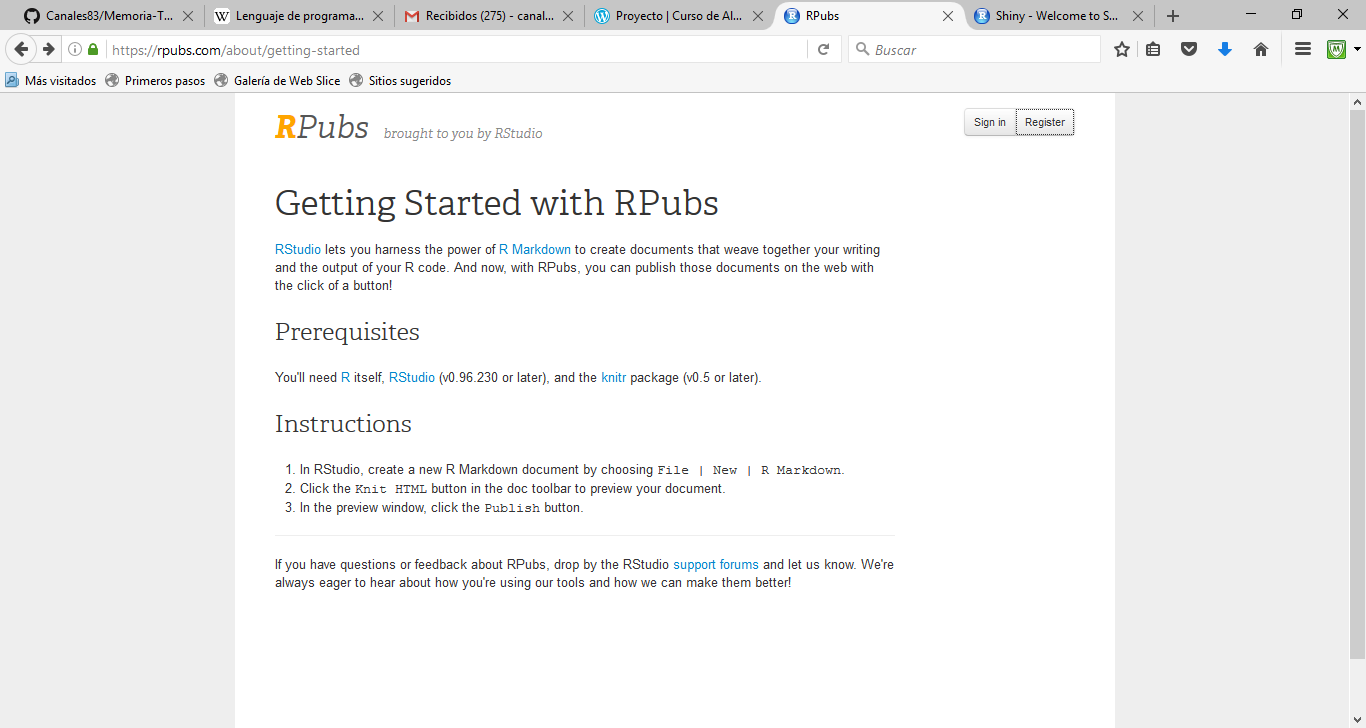
*RPubs* permite publicar los documentos creados en *RStudio* a través de *Markdown*, permite publicar el código y la salida en R en la web, de tal forma que sea accesible para cualquier usuario.

Se requiere que este instalado *R, RStudio* (v0.96.230 o posterior) y el paquete [*knitr*](https://github.com/yihui/knitr#readme) (v0.5 o posterior). Las instrucciones para comenzar a utilizarlo son:

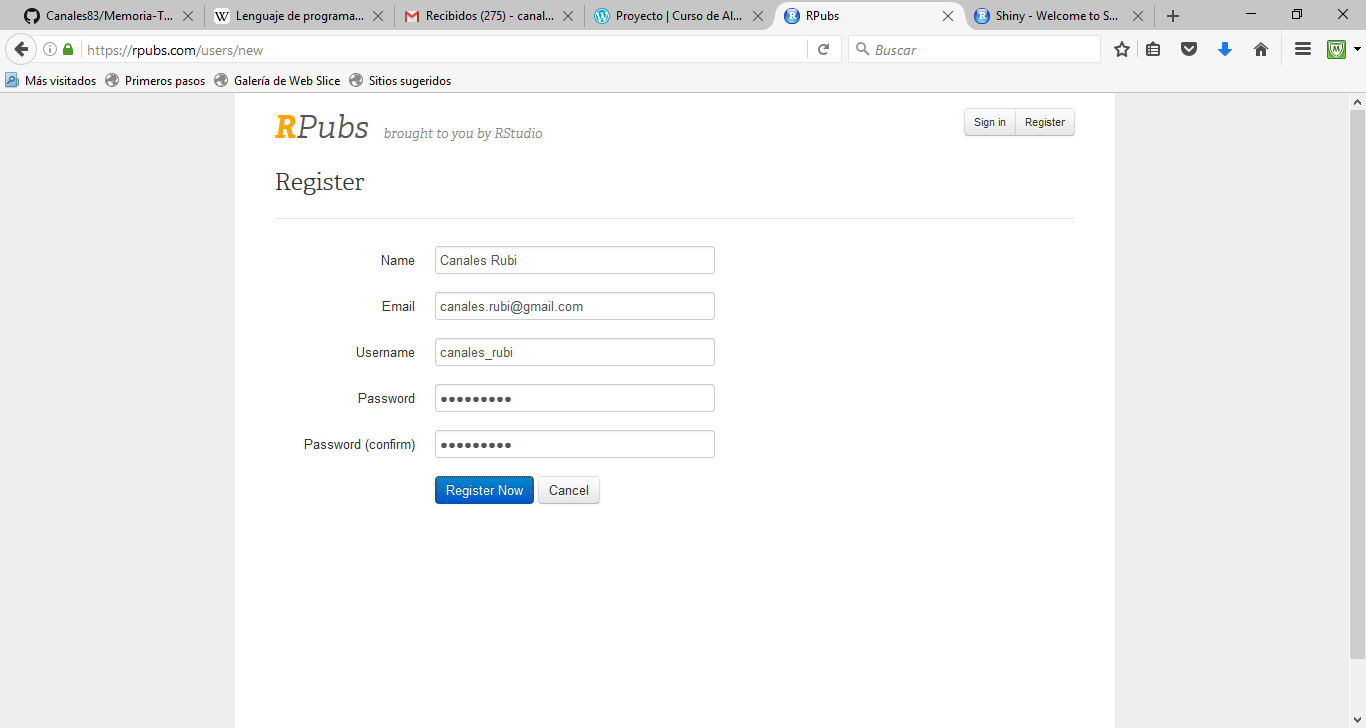
1. Accesar a la página:

<https://rpubs.com/about/getting-started>

1. En dicha página deberemos primeramente registrarnos, dar clic en el botón de *Register*:



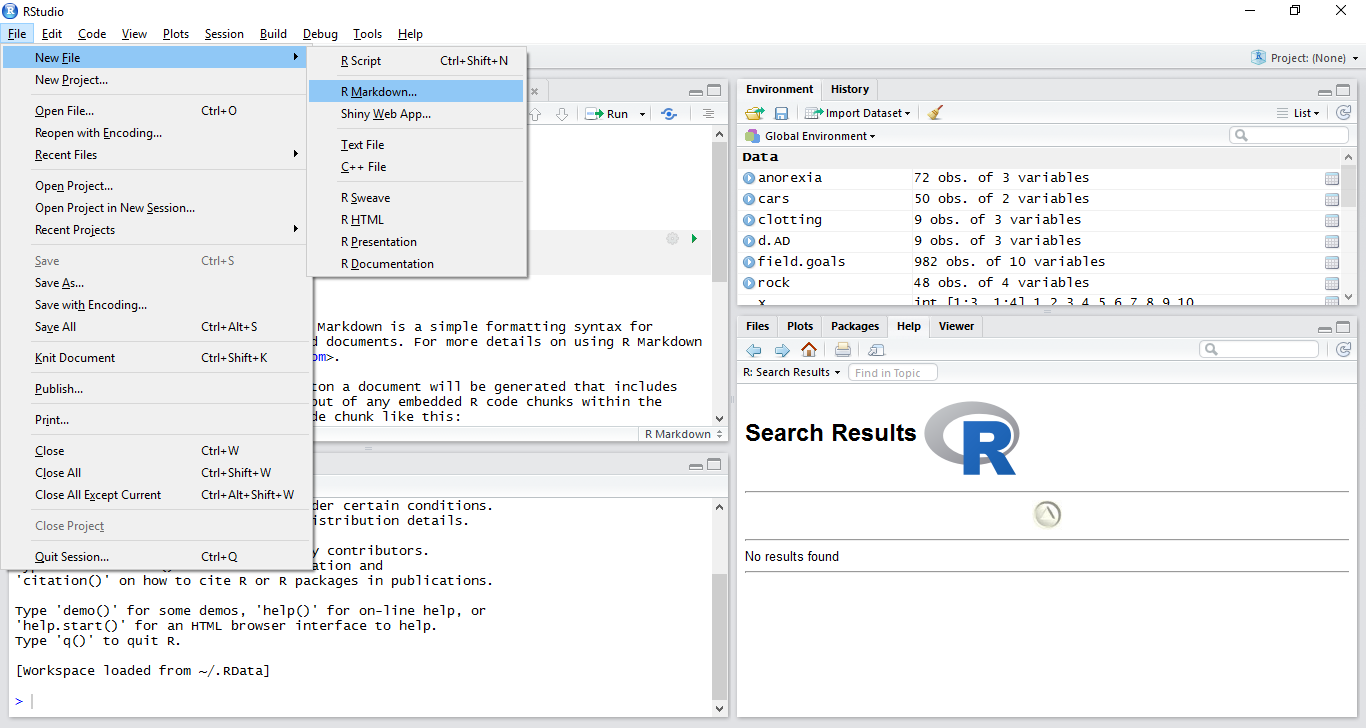
1. La siguiente página que nos mostrará será el registro en sí de una cuenta asociar y a través de la cual subiremos nuestros archivos para compartirlo. Nos solicitará nuestro nombre, email, usuario y *password*. Dar clic en *Registrer* *Now*, para darnos de alta:



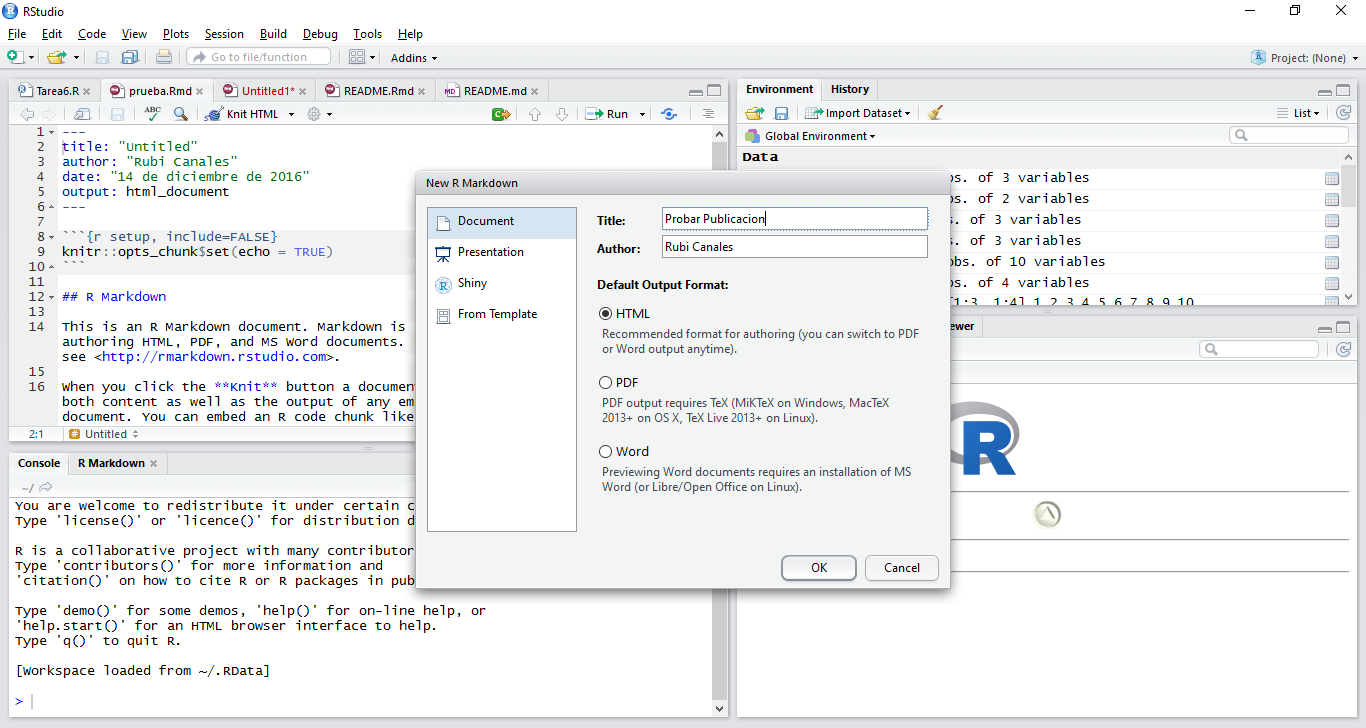
1. Una vez registrado nos enviará a la siguiente página:



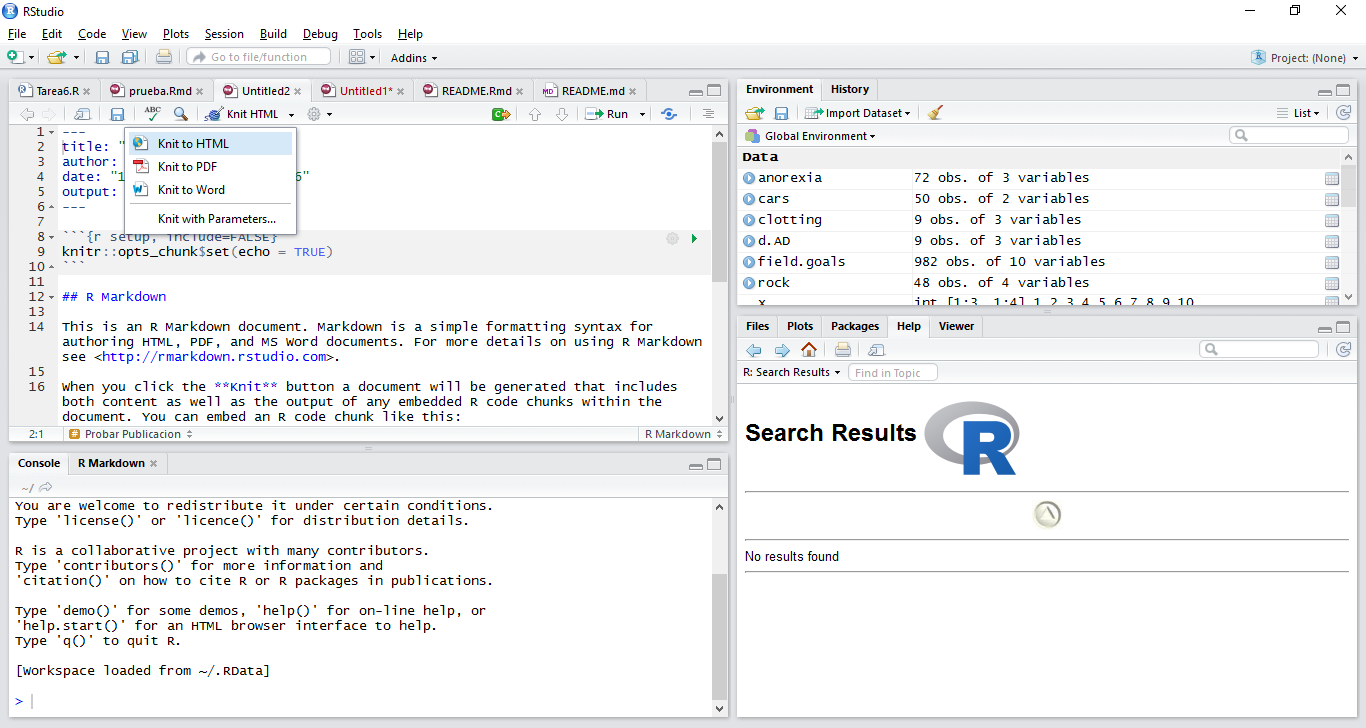
1. Dar clic en *Here’s how you get started*, para comenzar a trabajar en él.
2. En *RStudio*, crear un primer archivo de tipo R *Marktdown* por medio de *File🡪New file🡪 R Markdown:*



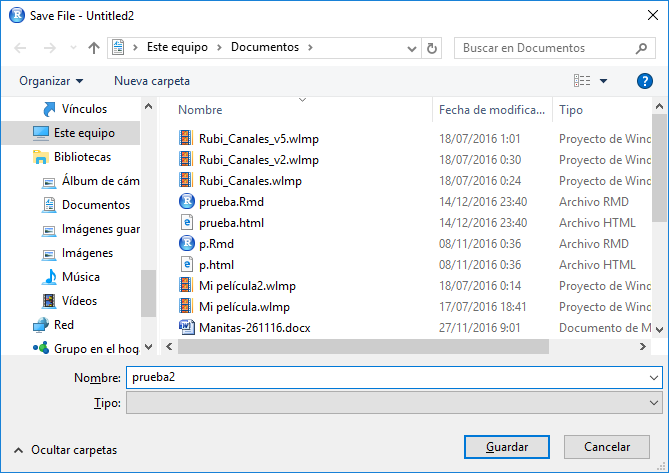
1. Nos pedirá un titulo y el tipo de archivo: html, pdf o word. Elegimos html:



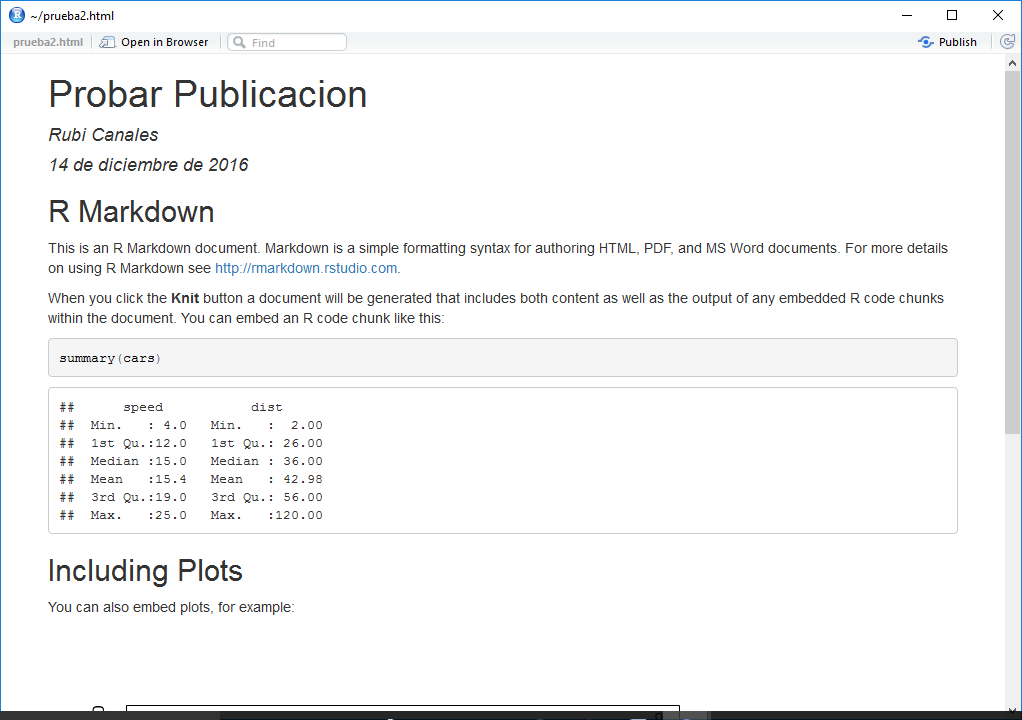
1. Generar el archivo html con ayuda de *Knit* HTML:



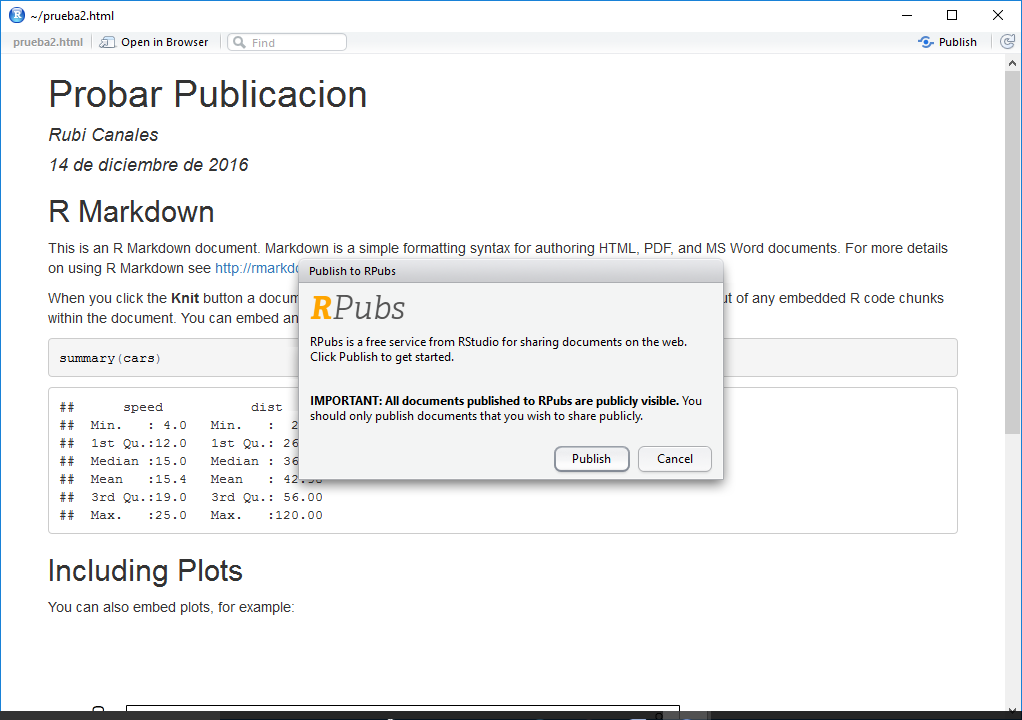
1. Guardar el archivo con un nombre y en la ruta deseada:



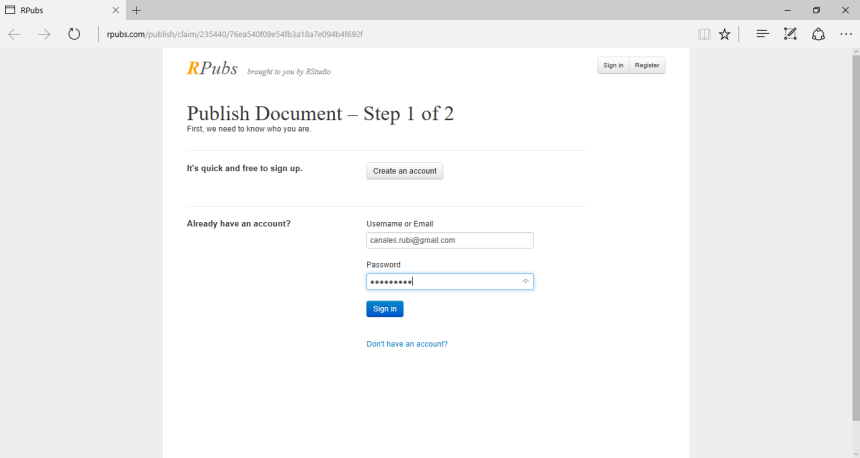
1. Una vez que se genere el html, dar clic en el botón superior derecho: *Publish*



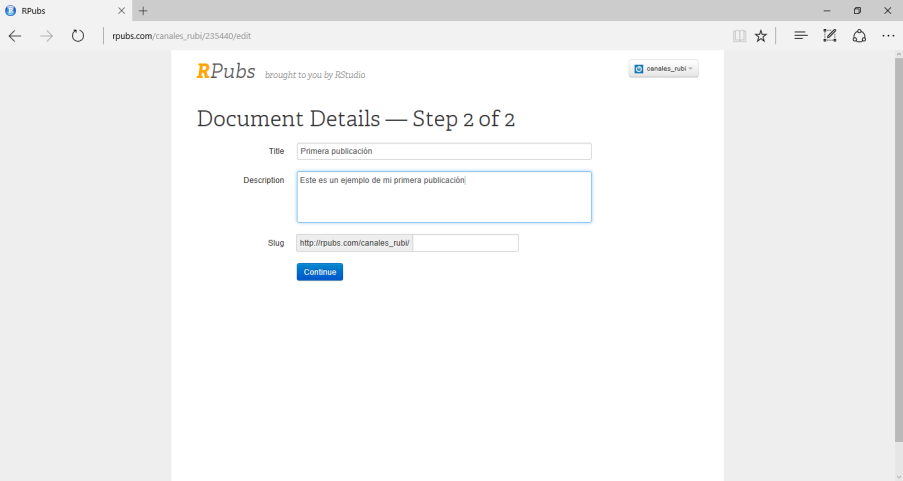
1. Nos saldrá un cuadro de dialogo , donde le indicaremos *Publish:*



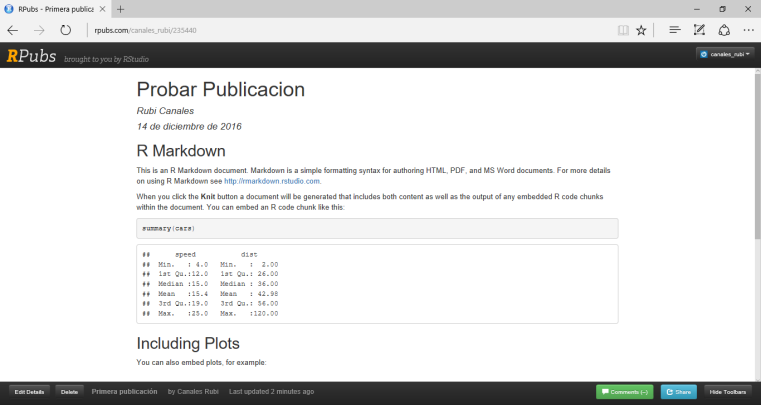
1. Si no estamos conectados en nuestra cuenta creada en el paso 1, deberemos conectarnos con nuestro user y *password*:



1. Colocar datos de la publicación que haremos, como el título y la descripción:



1. Y quedará publicado:



1. Si queremos revisar que códigos o documentos hemos publicado, iremos a la página principal de *RPubs* y ahí se mostrarán todos los archivos que hemos publicado con el nombre y descripción que pusimos en el paso 13 de este documento.



# Conclusiones

Con esta memoria técnica podemos ayudar en un futuro a instalar software necesario para las personas con perfil de científicos de datos como son: *R, R Studio, Git y GitHub, RPubs.*

*R* es un lenguaje de programación con enfoque estadístico, popular en la minería de datos, la investigación biomédica y las matemáticas financieras. Actúa como cualquier lenguaje de 4to generación puesto que contiene bibliotecas, palabras reservadas y sentencias cíclicas que permiten generar algoritmos que se pueden especializar estadísticamente.

Por otro lado *R Studio* es una herramienta que contiene una interfaz más amigable, familiarizándose con una interfaz más grafica que puede generar graficas más especializadas y más definidas.

Finalmente, tanto *Git* como *GitHub* son software que nos permite compartir archivos, en especial códigos fuentes, aunque también se pueden compartir documentos, de esta manera se hace uso del control de versiones para que varias o una persona pueda actualizar y guardar las versiones anteriores tantas veces lo requiera. Adicional, nos permitirá compartir información con cualquier persona a través de repositorios en internet.

Otra herramienta adicional para compartir archivos, códigos y ejecuciones de *R* es *RPubs* que nos permitirá compartir en la web códigos y ejecución de R a través de archivos tipo *Markdown*.

Esta memoria técnica tiene como principal características servir como guía para instalar y desinstalar estos software necesarios para cualquier proyecto que implica codificación de algoritmos.

# Bibliografía

[1] E. R. Rodríguez-Ávila, “*Curso de Almacenamiento de Datos y su Administración*”, Agosto-2016. [Online]. Available: https://eravilaipnada.files.wordpress.com/2016/08/ada2016b-temario-v1-00.pdf. [Accessed: 05-Nov-2016].

[2] Duncan Murdoch, “*R-3.3.2 for Windows*”, 2016. [Online]. Available: <https://cran.r-project.org/bin/windows/base/>. [Accessed: 10-Dic-2016].

[3] R Studio, “*R Studio*”, 2016. [Online]. Available: <https://www.rstudio.com/products/rstudio/>. [Accessed: 10-Dic-2016].

[4] Git member of software freedom conservancy, “*Git local branching on the cheap*”, [Online]. Available: <https://git-scm.com/downloads>. [Accessed: 10-Dic-2016].

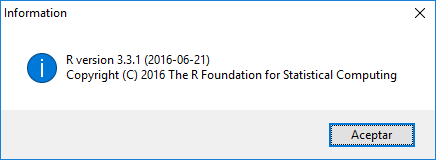
[5] GitHub, Inc, “*How people build software*”, 2016. [Online]. Available: <https://github.com/>. [Accessed: 10-Dic-2016].

[6] RStudio, “*Getting Started with RPubs*”. [Online]. Available: https://rpubs.com/about/getting-started. [Accessed: 15-Dic-2016].

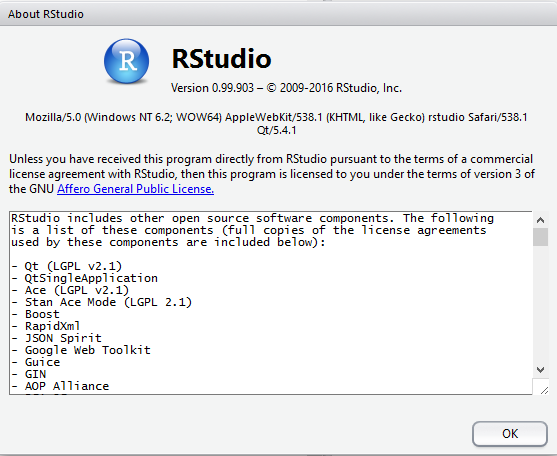
# Apéndice

* Plataforma tecnológica

*Software* R versión 3.3.1.



*RStudio*, versión 0.99.903.



*Git* versión 2.10

