

Clase 1: Introducción al Curso Análisis Avanzado de Datos II

Gabriel Sotomayor

2025-03-10



Presentación

Este curso profundiza en las técnicas de análisis multivariante, permitiendo examinar de manera integrada la interacción de múltiples factores en el estudio de problemas sociales. Se enfatiza la comprensión de los procedimientos y la interpretación rigurosa de los resultados, con un enfoque aplicado mediante el uso de herramientas computacionales, principalmente **R y RStudio**.

El curso prioriza la aplicación práctica de los métodos estadísticos sin requerir una profundización en sus fundamentos matemáticos. Al finalizar, el estudiante será capaz de analizar datos primarios y secundarios, identificar la técnica estadística adecuada según el contexto y generar informes que comuniquen de manera efectiva los hallazgos obtenidos.



Resultados de Aprendizaje

Resultado general: Desarrollar la capacidad de aplicar técnicas de estadística multivariante descriptiva e inferencial para analizar datos, formular hipótesis y construir modelos explicativos en investigaciones sociales, asegurando la correcta interpretación y comunicación de los resultados.



Resultados Específicos

1. Gestionar bases de datos complejas, asegurando su preparación y estructuración para el análisis.
2. Definir y estructurar problemas de análisis multivariante, formulando preguntas de investigación con sustento metodológico.
3. Seleccionar técnicas estadísticas adecuadas, justificando su aplicación en función del tipo de datos y los objetivos del análisis.
4. Utilizar herramientas computacionales especializadas, aplicando modelos en R y RStudio de manera efectiva.
5. Interpretar de forma rigurosa los resultados del análisis, identificando sus alcances y limitaciones.
6. Redactar informes técnicos y académicos, comunicando los hallazgos de manera clara, estructurada y fundamentada.



Contenidos

1. Gestión de Datos con R

- Introducción a la manipulación de datos con tidyverse.
- Cálculo e interpretación de estadísticos descriptivos.
- Visualización de información con ggplot2.
- Elementos introductorios de investigación reproducible.

2. Introducción al uso de muestras complejas en R

- Conceptos fundamentales de muestreo complejo.
- Diseño de encuestas y aplicación de ponderaciones.
- Inferencia estadística en muestras complejas.
- Manejo de encuestas con los paquetes survey y svyr.



Contenidos

3. Introducción a los Modelos Multivariados

- Rol de los modelos en las ciencias sociales.
- Diferencias entre enfoques exploratorios y confirmatorios.
- Repaso de conceptos estadísticos clave: covarianza, correlación e inferencia.
- Supuestos del análisis multivariante y su importancia.

4. Repaso de modelos de regresión

- Relevancia del control estadístico.
- Regresión lineal múltiple.
- Regresión logística binaria.



Contenidos

5. Análisis Factorial Exploratorio

- Aplicación en la investigación sociológica.
- Comparación entre análisis de componentes principales y factor común; supuestos del método.
- Métodos de extracción de factores, criterios de selección y técnicas de rotación.
- Interpretación de la matriz factorial y evaluación del ajuste. Cálculo de puntuaciones factoriales.
- Introducción al paquete psych.



Contenidos

6. Análisis Factorial Confirmatorio

- Diferencias principales con el análisis factorial exploratorio.
- Especificación e identificación del modelo, estimación de parámetros.
- Evaluación del ajuste, reespecificación y validez (convergente y discriminante).
- Ejemplo práctico de análisis confirmatorio.
- Introducción al paquete lavaan.



Contenidos

7. Análisis de Sendero

- Fundamentos y aplicación en ciencias sociales.
- Especificación del modelo de sendero y verificación de supuestos.
- Evaluación de la medición y capacidad confirmatoria.
- Ejemplo aplicado en investigación sociológica.

8. Modelos de Ecuaciones Estructurales

- Definición y aplicación de los SEM en ciencias sociales.
- Estructura del modelo y supuestos de la técnica.
- Estimación y evaluación del modelo, introducción de modificaciones.
- Ejemplos de modelamiento de ecuaciones estructurales.



Bibliografía

Ítem	Título	Autor	Año
1	Análisis Estadístico Multivariante - Un Enfoque Teórico y Práctico	De La Garza García, Jorge	2013
2	Análisis de datos multivariantes.	Peña, Daniel	2002
3	Análisis Multivariado Aplicado	Uriel Y Aldas	2005, 1ª Edición
4	Análisis Multivariante para las Ciencias Sociales	Levi J.P. y Varela J.	2001
5	Modelos de Ecuaciones Estructurales Cuadernos de estadística	Batista, J.M. y Coenders G.	2012
6	Introducción al Análisis de Regresión Lineal	Montgomery, Peck y Vining	2006
7	El análisis factorial como técnica de investigación en Psicología	Ferrando, P. J., &	2010



Ítem	Título	Autor	Año
		Anguiano- Carrasco, C.	



Bibliografía

Ítem	Título	Autor	Año
8	Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud.	Batista-Foguet, J.M., Coenders, G., & Alonso, J.	2004
9	El Path Analysis: conceptos básicos y ejemplos de aplicación.	Pérez, E., Medrano, L. A., & Rosas, J. S	2013
10	Modelos de ecuaciones estructurales	Ruiz, M.A	2010
11	RStudio para Estadística Descriptiva en Ciencias Sociales. Manual de apoyo docente para la asignatura Estadística Descriptiva	Boccardo, G. y Ruiz, F.	2019
12	R para Ciencia de Datos https://es.r4ds.hadley.nz/index.html	Wickham, H	2019
13	Exploring complex survey data analysis using R: A tidy introduction with {srvyr} and {survey}	Zimmer, S. A	2024



Evaluaciones

Evaluación	Fechas	Porcentaje
Tareas uso de R	24 de marzo 28 de abril 2 de junio 16 de junio	30%
Prueba	19 de mayo	35%
Trabajo final	1 de julio	35%



Ayudantías

El curso tiene dos ayudantes:

Fernanda Hurtado fernanda.hurtado@mail.udp.cl

Francisca Hernández francisca.hernandez_c@mail.udp.cl

Están disponibles para responder las dudas que puedan tener a lo largo del curso, tanto estadísticas como de uso de software.

Habrá sesiones de ayudantía cada 2 semanas aproximadamente, centradas en la aplicación de las técnicas que revisaremos en R.

También les acompañarán en la realización de tareas y trabajos de investigación.



Página del Curso

<https://aadii2025.netlify.app/>



Delegado de Curso

Las comunicaciones del curso con el equipo docente para temas colectivos deberán gestionarse de manera centralizada mediante un delegado, especialmente considerando que hay alumnos de distintas generaciones.

Esto es particularmente relevante para solicitudes respecto a evaluaciones.



Objetivos de la Sesión

Reflexionar sobre el uso de software estadístico en ciencias sociales. Introducir el uso de R y Rstudio para ciencias sociales.



¿Donde se sitúa la estadística en el proceso de investigación?



El uso de software y estadística es una pequeña parte del proceso de diseño de una investigación, por lo que es importante no confundir ambos procesos ni anteponer el uso de determinadas herramientas al proceso general de investigación, debe primar el problema a investigar



USO DE SOFTWARE EN CIENCIAS SOCIALES

Progresivamente se ha generalizado el uso de software estadístico en ciencias sociales, abriendo grandes posibilidades de realizar análisis más complejos y facilitando su uso.

Riesgo de falta de formación estadística y reflexividad (calidad del análisis)

Dos niveles de manejo del software:

1. Ejecución básica y correcta interpretación de las salidas estadísticas (necesario mínimo para utilizar la herramienta e interpretar otras investigaciones).
2. Manejo avanzado de software y procedimientos estadísticos. Nivel deseable para un uso reflexivo de las herramientas y lectura crítica de otras investigaciones.



¿QUÉ SOFTWARE EXISTEN Y CUÁL UTILIZAR?



¿QUÉ SOFTWARE EXISTEN Y CUÁL UTILIZAR?

Dimensión / Lenguaje	R	Python	SPSS	Excel	Stata
Alcance	General, orientación multidisciplinar	General, orientación multidisciplinar	Limitado, orientado a Ciencias Sociales	Limitado, orientado a administración	Limitado, orientado a Economía
Licencia	Libre (freeware)	Libre (freeware)	Pagada (versión de prueba limitada)	Pagada (versión de prueba limitada)	Pagada (versión de prueba limitada)
Aprendizaje	Sintaxis, poco intuitivo	Sintaxis, poco intuitivo	Botones y sintaxis, intuitivo	Botones y sintaxis, intuitivo	Botones y sintaxis, intuitivo
Visualización	Avanzada	Intermedia	Básica	Intermedia	Intermedia
Análisis de texto	Intermedio, poca eficiencia	Avanzado, amplia eficiencia	No	No	No
Minería Datos	Intermedio, poca eficiencia	Avanzado, amplia eficiencia	No	No	No
Sistema operativo	Windows, Mac OS, Linux	Windows, Mac OS, Linux	Windows, Mac OS	Windows, Mac OS	Windows, Mac OS

(Boccardo y Ruiz, 2018)



¿QUÉ ES R?

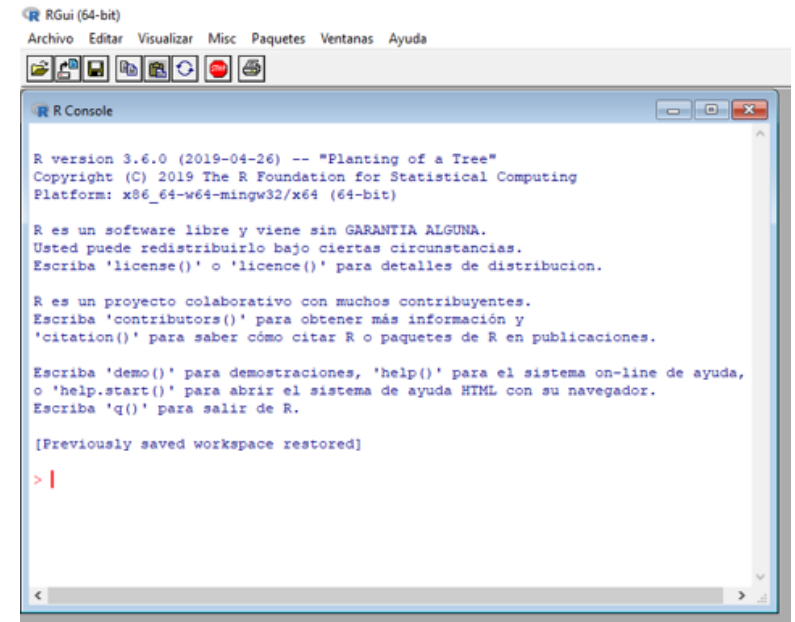
R es un software y un lenguaje de programación, ambos de carácter libre, enfocado principalmente en el análisis y visualización de datos. Es desarrollado por The R Foundation for Statistical Computing.

Funciona con una lógica de objetos, sobre los que podemos trabajar con operadores y funciones.

El hecho de que sea un lenguaje de programación nos permite programar funciones, por lo que es posible pasar de ser usuarios a programadores.

Instalamos una versión básica y podemos agregar paquetes con nuevas funciones.

Su interfaz es poco amigable, la solución: R Studio.



```
RGui (64-bit)
Archivo  Editar  Visualizar  Misc  Paquetes  Ventanas  Ayuda

R Console

R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

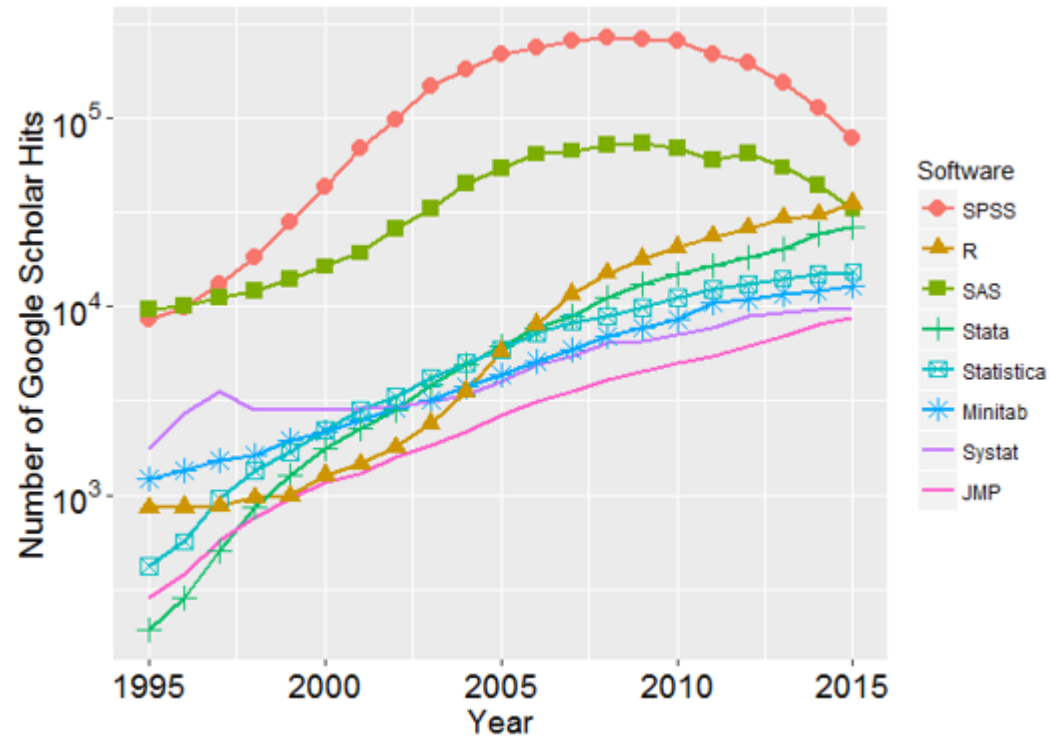
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

[Previously saved workspace restored]

> |
```



¿PORQUÉ APRENDER Y USAR R Y RSTUDIO?



¿PORQUÉ USAR R?

1. Es libre y gratis.
2. Por lo anterior ofrece grandes posibilidades de personalización de los análisis y amplitud de herramientas (en actualización).
3. Amplia y creciente comunidad de usuarios desarrollando funciones y resolviendo problemas en internet (la mayoría en inglés).
4. Su sintaxis es simple e intuitiva.



¿PORQUÉ TRABAJAR CON SINTAXIS?

1. **Replicabilidad:** Permite a otros entender como construimos nuestros resultados y replicarlos, cada vez más importante en la ciencia.
2. **Eficiencia:** En condiciones “reales” de trabajo resulta más eficiente que trabajar con botones (aunque ahora pueda parecer contraintuitivo)
3. **Control:** Permite tener claridad de todas las etapas de nuestros análisis, cuestión que en otros software son realizadas por defecto.

(Elousa, 2009 en Boccardo y Ruiz, 2018: 8-9)



R STUDIO

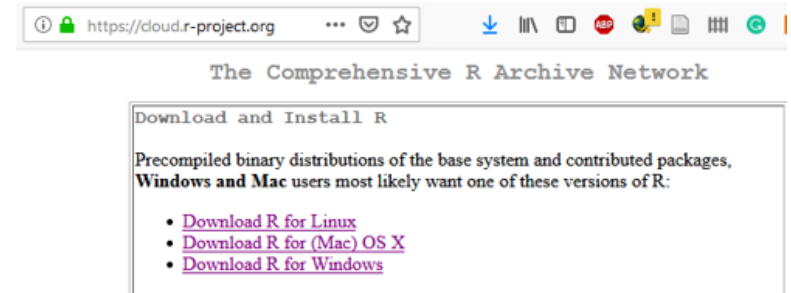


INSTALAR R Y RSTUDIO

Instalar R

Podemos encontrar R para Windows, Linux y Mac en <https://cloud.r-project.org/>

Debemos descargar el paquete base, luego ejecutar el instalador, e instalarlo según las opciones predeterminadas



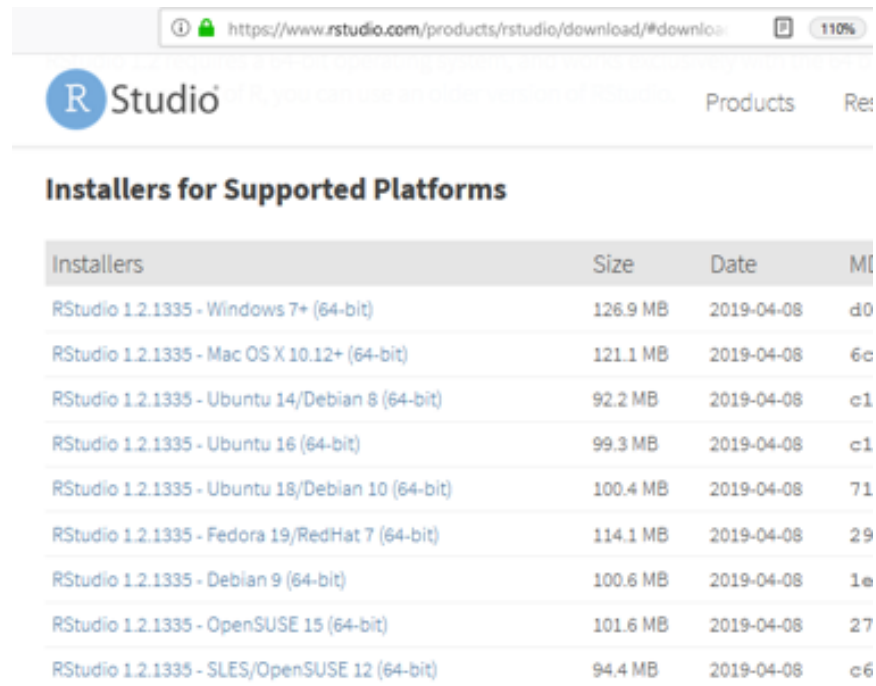
INSTALAR R Y RSTUDIO

Instalar R Studio

Podemos encontrar R Studio en su sitio web, y bajar la versión gratis, que contiene todas las funciones <https://www.rstudio.com/products/rstudio/download/>

Debemos fijarnos en elegir el sistema operativo correcto.

Igualmente lo ejecutamos e instalamos según las opciones predeterminadas

A screenshot of the RStudio website's download page. The browser address bar shows the URL 'https://www.rstudio.com/products/rstudio/download/#download'. The page features the RStudio logo and navigation links for 'Products' and 'Resources'. Below the header, the section 'Installers for Supported Platforms' is displayed, containing a table with columns for 'Installers', 'Size', 'Date', and 'Mirror'. The table lists various operating systems and their corresponding installer sizes and dates.

Installers	Size	Date	Mirror
RStudio 1.2.1335 - Windows 7+ (64-bit)	126.9 MB	2019-04-08	d0
RStudio 1.2.1335 - Mac OS X 10.12+ (64-bit)	121.1 MB	2019-04-08	6c
RStudio 1.2.1335 - Ubuntu 14/Debian 8 (64-bit)	92.2 MB	2019-04-08	c1
RStudio 1.2.1335 - Ubuntu 16 (64-bit)	99.3 MB	2019-04-08	c1
RStudio 1.2.1335 - Ubuntu 18/Debian 10 (64-bit)	100.4 MB	2019-04-08	71
RStudio 1.2.1335 - Fedora 19/RedHat 7 (64-bit)	114.1 MB	2019-04-08	29
RStudio 1.2.1335 - Debian 9 (64-bit)	100.6 MB	2019-04-08	1e
RStudio 1.2.1335 - OpenSUSE 15 (64-bit)	101.6 MB	2019-04-08	27
RStudio 1.2.1335 - SLES/OpenSUSE 12 (64-bit)	94.4 MB	2019-04-08	c6



EL PROBLEMA DE JAVA X64

Para algunos análisis R se basa en Java, el problema es que nuestros computadores suelen instalar de manera automática la versión de 32 bits, porque los navegadores de internet funcionan (o funcionaban) en esta arquitectura, pero nuestros computadores y R funcionan a 64 bits.

Para resolverlo debemos descargar e instalar la versión offline de 64 bits de Java
<https://www.java.com/es/download/manual.jsp>



PROGRAMACIÓN POR OBJETOS Y FUNCIONES

Tipos de estructuras de datos en R

Vector (vector): columna o fila de datos de un mismo tipo (una variable individual)

Listas (list): Nos permiten agrupar objetos de distinto tipo.

Matrices (matrix): arreglo de dos dimensiones de datos de un mismo tipo (conjunto de variables)

Data.frame (base de datos): Matriz de datos en el que las columnas tienen asignado nombres, y que permite usar todo tipo de datos.



PROGRAMACIÓN POR OBJETOS Y FUNCIONES

Tipos de variables en R

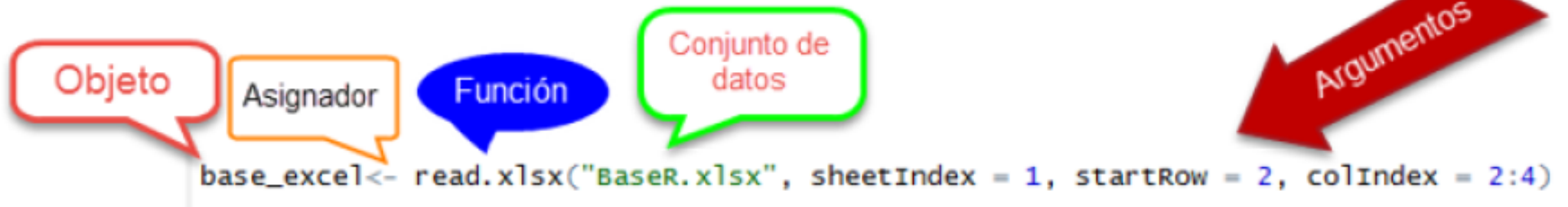
Numérico (numeric; números con decimales)

Números enteros (integer)

Lógicos: Verdadero o Falso

Variables cadena o texto (string / character)

Factores (factor): para variables cualitativas (nominales)



¿COMO CONTINUAR?

Dominar R es un camino de mediano plazo y es importante desarrollar habilidad de “hacking”, es decir:

1. estar dispuesto a buscar respuestas de manera autónoma
2. saber donde buscar la información
3. utilizar nuevos datos y paquetes por su cuenta

Donde buscar:

- GOOGLE! (preferentemente en inglés) y youtube
- Pedir ayuda a otros usuarios
- Pestaña “help”
- <https://www.rdocumentation.org/>
- <https://stackoverflow.com/>
- Chat GPT y otros LLM (Gemini, Claude, Deepseek, Grok, etc)*



BIBLIOGRAFÍA Y OTROS RECURSOS PARA PROFUNDIZAR

Boccardo, Giorgio y Ruiz, Felipe. Uso de RStudio para Estadística Univariada en Ciencias Sociales. Manual de apoyo docente para la asignatura Estadística Descriptiva. Santiago: Departamento de Sociología, Facultad de Ciencias Sociales, Universidad de Chile.
[Disponible aquí.](#)

Grolemund, Garrett y Wickham Hadley. R para Ciencia de Datos. <https://es.r4ds.hadley.nz/>

Paradis, E. (2003). R para Principiantes. Francia: Institut des Sciences de l'Évolution.
Disponible en: https://cran.r-project.org/doc/contrib/rdebuts_es.pdf

Sitios web útiles

<http://www.sthda.com>

Cursos online (inglés)

Especialización en Ciencia de Datos Universidad John Hopkins
<https://www.coursera.org/specializations/jhu-data-science>

Data Camp <https://www.datacamp.com/courses/tech:r>

