

**PLAN FEDERAL DE FORMACIÓN Y CAPACITACIÓN EN GESTIÓN PÚBLICA 2022**  
**DIRECCIÓN DE CAPACITACIÓN FEDERAL Y DIFUSIÓN VIRTUAL - INAP****1. Origen de la demanda:**

Instituto Nacional de la Administración Pública (INAP)

**2. Prestador de la actividad:**

Universidad Nacional de Tres de Febrero – Maestría en generación y análisis de información estadística

**3. Nombre de la actividad**

Taller para el análisis de datos en la administración pública: Introducción al uso de R y R-studio

**4. Fundamentación**

La implementación de políticas públicas, cualquiera sea la temática abordada y el nivel de ejecución alcanzado requieren de información cuantitativa para su fundamentación, desarrollo de las propuestas, seguimientos, evaluación de resultados e impactos obtenidos.

Esta información es generada a partir relevamientos y/o rescate de datos de registro utilizando el herramental provisto por la Ciencia Estadística, tanto en lo que respecta a la generación como al análisis de los datos obtenidos que se transformaran en la información de referencia para el logro de las metas y objetivos propuestos que dan sentido a la

planificación y gestión de gobierno.

Este curso propone el abordaje de algunas de estas herramientas de Estadística, haciendo principal foco en los procesos de construcción, análisis y visualización de la información de encuestas o registros a través del software R de código abierto para fortalecer el perfil profesional de los agentes de la administración pública.

En este marco la Secretaría de Gestión y Empleo Público y la Universidad Nacional de Tres de Febrero firmaron un convenio de cooperación a partir del cual se formula esta actividad de capacitación con el objetivo de desarrollar a nivel federal una mejora en las capacidades de los equipos para lograr los objetivos y resultados de gestión propuestos.

## 5. Contribución esperada

Nuestras vidas se encuentran atravesadas por la generación de datos: votar, tomar un transporte público, responder una encuesta o el relevamiento de información en la aplicación de una política pública. Por este motivo, implementar políticas públicas es también manejar con información para tomar decisiones y evaluar resultados. En este contexto se convierte en una necesidad el desarrollo de perfiles que conozcan y utilicen herramientas para procesar, transformar, analizar y visualizar información.

R es un lenguaje de código abierto que resulta amigable para quienes no tienen experiencia previa en programación. Nos brinda herramientas integradas que permiten consolidar un flujo de trabajo articulado en un mismo software.

**6. Eje formativo del Plan Federal en el que se inscribe:**

Eje de formación instrumental: Consiste en el desarrollo de módulos temáticos sobre herramientas concretas para la gestión cotidiana de las y los trabajadores de nivel local.

**7. Perfil del/la participante y requisitos (público objetivo)**

Trabajadores y trabajadoras de nivel operativo de las administraciones públicas provinciales y municipales que se procesen información y/o quieran incorporar conocimientos de procesamiento de datos.

El curso no requiere conocimientos previos.

**8. Objetivos**

Capacitar a los cursantes en la transformación de la información para el seguimiento y diseño de políticas públicas.

Introducir al uso de software libre para el análisis de datos en la administración pública.

Acompañar en el uso del lenguaje de programación R para fortalecer el perfil profesional en el procesamiento, análisis y visualización de información, de los profesionales de la administración pública.

Que los cursantes puedan transformar y modelar datos a través de la utilización de paquetes y funciones.

Que los asistentes puedan desarrollar visualizaciones que fortalezcan la comunicación de las políticas públicas, a través de aprender la estructura básica de una representación gráfica en R.

## 9. Contenidos

### UNIDAD I: Introducción al procesamiento de datos

En esta unidad realizaremos una introducción al problema de los datos y la información en la administración pública. ¿Por qué es tan importante obtener datos y producir información? Realizaremos un recorrido de las diferentes herramientas que se utilizan en diferentes organismos del estado. Destacaremos la oportunidad que se nos presenta al utilizar herramientas gratuitas y basadas en software libre vs las herramientas propietarias. En particular, el uso de R y Rstudio para procesar datos y construir información confiable. Guiaremos a los participantes en la instalación del software, conoceremos la interfaz de estas herramientas y comenzaremos a aplicar los comandos básicos.

Ejercicios de aplicación: Mostraremos ejemplos para cargar diversos conjuntos de datos, relacionados a diferentes problemáticas de interés y acción públicas, para realizar las primeras métricas.

## **UNIDAD II:** Transformación de datos con R

En esta unidad veremos cómo leer los diferentes archivos con los que habitualmente trabajamos en la administración pública (excel, csv, sav, etc). Con estos datos realizaremos diferentes transformaciones que nos permitan construir las métricas que necesitamos. Vamos a filtrar datos por alguna característica de interés, seleccionaremos las variables que nos interesan analizar, vamos a producir agrupaciones y sacar métricas por grupos (cantidad, proporciones, etc).

Ejercicios de aplicación: Conoceremos una biblioteca de R de producción nacional **EPH** para trabajar con las bases de datos la EPH (Encuesta permanente de hogares), encuesta desde la cual se obtienen estadísticas sociodemográficas como tasa de desocupación, población económicamente activa, pobreza, etc.

## **UNIDAD III:** Visualizando de información mediante gráficos

Una vez que obtenemos la información que necesitamos, un gran desafío es comunicar los resultados obtenidos. Esto se torna especialmente crítico en la administración pública, donde necesitamos comunicar los resultados a nuestros colegas, otros organismos, decisores de políticas públicas y en muchas circunstancias al público general.

Por este motivo, veremos la forma de construir gráficos R. Discutiremos cómo elegir el gráfico indicado de acuerdo a la información a presentar. Veremos gráficos de barra, de columna, de líneas, boxplot. Conoceremos la principal librería de R para extender nuestras posibilidades de construcción de gráficos.

Ejercicios de aplicación: Construiremos gráficos para presentar la información obtenida para un conjunto de datos, que estos sean significativos y claros.

#### **UNIDAD IV:** Introducción al análisis de información geográfica

En muchos casos en diferentes organizaciones públicas contamos con información que puede/debe analizarse desde una dimensión y perspectiva territorial. Las problemáticas pueden variar de provincia en provincia, departamento a departamento dentro de una provincia y barrio a barrio en un municipio. Por lo tanto en esta unidad presentamos una serie de recursos que R nos provee para leer este tipo de información, transformarlo y visualizarlo cuando sea necesario y oportuno.

Ejercicios de aplicación: Desarrollo de nuestros primeros mapas en R para analizar una problema público desde una perspectiva territorial.

#### **UNIDAD V:** Presentación de informes y documentos con R

Comunicar es clave en todos los niveles del estado para todos los organismos. Hay que comunicar para tomar decisiones, para informar a un organismo o al público. Para eso en general construimos informes, documentos o tableros, en donde insertamos gráficos, imágenes, análisis

de los resultados, etc. En R contamos con herramientas que nos permiten realizar estos informes sin recurrir al Word o al Power Point. Dentro de R podemos construir flujos de trabajo que impliquen la lectura de datos, la transformación, obtención de métricas, visualizaciones y la construcción del informe. A esto llamamos flujo integrado de trabajo en R.

Ejercicios de aplicación: Sistematizar las distintas etapas del flujo de trabajo en el análisis de datos en informes. Que permitan al analista de información compartir su trabajo con otros sectores de la administración pública y con diferentes tomadores de decisión.

#### **10. Estrategias metodológicas y recursos didácticos**

La cursada se organiza en 5 unidades secuenciales. Se promueve el acompañamiento de los participantes a través de foros y envío de materiales complementarios. El curso consta de cinco guías que abordan cada uno de los tópicos desarrollados en los contenidos para su seguimiento a través de material de lectura y distintos recursos audiovisuales (videos explicativos para cada módulo y clases grabadas) y resolución de guías para acompañar la introducción en el uso de R.

#### **11. Descripción de la modalidad y cantidad de horas**

El curso será dictado bajo la modalidad virtual tutorada bajo instancias asincrónicas y dos encuentros opcionales sincrónicos (la grabación se encontrará disponible en la plataforma) las que se estiman en un total de 24 horas.

#### **12. Bibliografía para el/la participante**

Antonio Vázquez Brust (2015). “Ciencia de datos para gente sociable: Una introducción a la exploración, análisis y visualización de datos”. Disponible en línea, visitado por última vez el 17/04/2022: [https://bitsandbricks.github.io/ciencia\\_de\\_datos\\_gente\\_sociable/](https://bitsandbricks.github.io/ciencia_de_datos_gente_sociable/)

Garrett Grolemund (2014). “Hands-On Programming with R”. O’reilly. Disponible en línea, visitado por última vez el 17/04/2022: <https://rstudio-education.github.io/hopr/>

Hadley Wickham, Garrett Grolemund (2014). “R for data science: import, tidy, transform, visualize and model data”. O’reilly. Disponible en línea, visitado por última vez el 17/04/2022: <https://es.r4ds.hadley.nz/>

Pablo Casas (2015). “Libro vivo de Ciencia de Datos”. Disponible en línea, visitado por última vez el 17/04/2022: <https://librovivodecienciadedatos.ai/>

### **13. Evaluación de los aprendizajes y de la actividad**

Se realizará un seguimiento general de la participación de los y las agentes inscriptos en la actividad, a través de la participación en las actividades asincrónicas y sincrónicas (opcionales) con grabación disponible en plataforma.

Como evaluación individual, se realizará un trabajo práctico final integrador utilizando R para analizar un conjunto de datos de interés del cursante.

### **14. Requisitos de asistencia, aprobación y certificación**

Los participantes de cursos tutorados virtuales del INAP serán calificados como:



- Ausente
- Libre
- Presente
- Aprobado
- Desaprobado

Las formas de participación en el entorno virtual de aprendizaje serán las siguientes:

- Lecturas de los textos y visualización de las exposiciones del docente.
- Participación en foros.
- Realización de la guía de actividades prácticas (ejercicios autoasistidos en formato de control de lectura).
- Aprobación del trabajo práctico final integrador.

Para recibir el certificado de aprobación: Cumplimiento de al menos el 80% de las actividades prácticas de seguimiento propuestas (lectura y visualización de las exposiciones docentes, participación en foros y entrega de la guía de actividades prácticas) más la aprobación del trabajo final para el certificado de aprobación.

Para recibir el certificado de asistencia: Cumplimiento de al menos el 80% de las actividades prácticas de seguimiento.

### 15. Perfil e información de la/el docente

Flavio Scargiali. Sociólogo de la Universidad de Buenos Aires. Maestrando en Generación y Análisis de Información Estadística por la Universidad Nacional de Tres de Febrero. Se

especializa en el análisis de datos globales y analytics. Actualmente se desempeña como analista en Equifax S.A. Se ha desempeñado como docente en cursos de introducción al lenguaje R en diferentes universidades (UNTREF, FSOC UBA).

Leandro Colisko. Licenciado en Ciencia Política Universidad Nacional de La Matanza. Maestrando en Generación y Análisis de Información Estadística por la Universidad Nacional de Tres de Febrero. Cuenta con una amplia experiencia como analista de datos estadísticos. Actualmente se desempeña en el Departamento de Estadísticas en la Empresa Argentina de Navegación Aérea. Se ha desempeñado como docente en cursos de introducción al lenguaje R en diferentes universidades (UNTREF, FSOC UBA).

## 16. Duración

24 horas distribuidas en 7 semanas

Semana	Tipo de actividad	Contenido: módulo temático	Evaluación
Semanas 1	<u>Asincrónico</u> (4 horas)	Presentación Unidad I: Introducción a R. ->Ejercicios de aplicación	-
Semana 2	<u>Asincrónico</u> (4 horas)	Unidad II: Introducción a Tidyverse y lectura de datos. ->Ejercicios de aplicación	-

<b>Semana 3</b>	<u>Sincrónico</u> 2 horas (opcional. La grabación quedará disponible) y Asincrónico 2 horas.	Unidad III: Dplyr, análisis exploratorio de datos. ->Ejercicios de aplicación	-
<b>Semana 4</b>	<u>Asincrónico</u> (4 horas)	Unidad IV: Visualización de datos con ggplot2. ->Ejercicios de aplicación	-
<b>Semana 5</b>	<u>Asincrónico</u> (4 horas)	Unidad V: Introducción a datos geográficos con R. ->Ejercicios de aplicación	-
<b>Semana 6</b>	<u>Sincrónico</u> 2 horas (opcional. La grabación quedará disponible)	Unidad IV: Creación de informes con Rmarkdown. ->Ejercicios de aplicación	-
<b>Semana 7</b>	Asincrónico (2 horas).	Consultas y dudas	Trabajo Práctico Final.

## 17. Lugar

Plataforma de la Universidad.