

# Proyecto 1

Rodrigo Zepeda

Otoño 2021

Por equipo a lo más de 3 o de manera individual. Ustedes pueden divorciarse o juntarse como quieran y cuando quieran. No existen equipos de 4 (un equipo de 4 se divide en dos equipos de dos)

## PROYECTO ALTERNATIVO 2021

Equipos de mujeres que participen en el concurso *Mujeres en el territorio* en la parte de visualización de datos <https://sistemas.sedatu.gob.mx/MET> . Si eligen hacer esto nos regimos por las reglas del concurso y lo de abajo no aplica (más allá de que los equipos independientemente de la convocatoria son de máximo 3 para que valga para clase).

Si alguien consigue otro análisis de datos que sea un concurso o un proyecto de su trabajo con gusto lo discutimos para que ese sea su proyecto.

## Objetivo

El proyecto consiste en realizar análisis de alguna base de datos (**que sea un CENSO no una MUESTRA**) de la vida real e incluir la interpretación de lo que salga. Algunas opciones de bases pueden encontrarse en: [el portal de datos de CDMX](#), [el portal de datos del Gobierno Federal](#), [el portal de datos de Jalisco](#), [el portal de datos de Nuevo León](#), [el portal de datos de Colombia](#), [la fundación para la Ciencia Abierta](#), [Kaggle](#). La base elegida debe estar disponible para que el profesor pueda replicar el análisis ya sea en un sitio web o mandándosela directamente al correo.

**NOTA** Si hicieron algún cambio en excel a la base de tal forma que no sea exactamente la misma que está en el sitio web es necesario compartir *la que ustedes usaron* no la de la web.

Específicamente el proyecto debe incluir al menos 10 gráficas de tipo distinto (es decir si ya hiciste un histograma no cuenta hacer *otro* histograma aunque sea de otra variable). Así mismo el proyecto debe incluir el uso de al menos 5 estadísticos vistos en clase u obtenidos de otra fuente (ej. la media, mediana, varianza, etc). Finalmente deben usarse al menos una vez cada una de las siguientes funciones de tidyverse: `filter`, `summarise`, `mutate`, `select`, `group_by`.

Se sugiere evitar datos geoespaciales por la dificultad del análisis. Aunque si te interesa puedes revisar [este libro](#).

## Fecha sugerida de entrega

20 de septiembre Al correo del profesor: [rodrigo.zepeda@itam.mx](mailto:rodrigo.zepeda@itam.mx) con copia a [rzepeda17@gmail.com](mailto:rzepeda17@gmail.com)

## Indicaciones

1. El proyecto debe ser entregado en pdf o html.
2. El proyecto debe incluir una descripción breve de qué se trata la base y de dónde se obtuvo. Ejemplo “es una base de los registros de interrupciones legales del embarazo en CDMX la cual se obtuvo del siguiente link: <https://datos.cdmx.gob.mx/dataset/interrupcion-legal-del-embarazo>”

3. Para las gráficas incluir una interpretación. Pueden hacerse las gráficas vistas en clase así como otras gráficas (sugerencia: ver [este link](#)). Si se desea se puede utilizar la [librería meme](#) y que una gráfica sea un meme sin interpretación. El *caption* de la gráfica debe incluir al autor de la misma. Si el equipo es de dos personas se deben hacer 5 y 5 gráficas. Si es de 3 personas se deben hacer 3, 3 y 3. La 4a gráfica puede ser de quien sea o ser un meme.
4. Para que una gráfica sea buena debe cumplir con los siguientes criterios:
  - a. *Reproducibilidad* cualquiera con la base y el código puede reproducir la gráfica.
  - b. *Mismo idioma* todo lo que está en la gráfica debe aparecer en el mismo idioma (nada de graficar la variable en el eje x diciendo AGE y en el eje y diciendo PERSONAS).
  - c. *Incluir conceptos no variables* los ejes y títulos de la gráfica deben incluir conceptos (por ejemplo **Peso de una persona (kg)**) y no variables (por ejemplo `dataset$wt`).
  - d. *Sin errores ortográficos*.
  - e. *La gráfica habla por sí misma* La gráfica podría extraerse del texto y debería, por sí misma, comunicar toda la información incluyendo fuente, variables a graficar, periodo temporal, significado de los datos. Piensa en que si alguien va a poner tu gráfica sin ningún otro contexto en redes sociales debe entenderse qué es.
  - f. *Escalas en notación humana* Las escalas no deben de tener notación científica del estilo 1.e7. Se sugiere usar la librería `scales`.
  - g. *Ordenamiento de categorías* Las variables categóricas deben seguir un orden ya sea alfabético o de mayor a menor. Se sugiere usar `factor`.
  - h. *Gráfica correcta para los datos* No realizar gráficas de variables continuas para datos discretos, por ejemplo.
6. En el mismo documento o en un `Rscript` (.R) aparte, incluir todo el código a realizarse y comentarios a dicho código. El código debe seguir las [recomendaciones de tidyverse](#).
7. No se consideran proyectos donde el código no corra en mi computadora. No se consideran proyectos donde el equipo no pueda responder preguntas del tema en caso de que se requieran. Ni se consideran proyectos cuyo código sea demasiado similar (25%) al de semestres pasados.
8. Al final del proyecto debe incluirse qué hizo cada quien desglosado por los siguientes apartados: **Equipos de una persona pueden ahorrarse esto porque lo hicieron todo solos.**
  - a. Obtención de la base. *Quién se encargó de obtener la base que se iba a usar*
  - b. Determinación del objetivo. *El proyecto debe incluir un objetivo usualmente el objetivo es convencer al lector de algo: convencer que vamos mejor/peor económicamente, convencer que en X ó Y zona hay más clientes de una empresa o convencer de que fumar está asociado a cáncer. Ustedes deciden el objetivo.*
  - c. Descripción de la base. *Quién describió la base en el reporte*
  - d. Código en R para los estadísticos (por estadístico). *Quién realizó el cálculo de los estadísticos (por ejemplo quién calculó la media y quién la varianza).*
  - e. Interpretación de los estadísticos (por estadístico). *Quién se encargó de traducir ese número en palabras y explicar qué significa en el reporte.*
  - f. Código en R que usa `mutate`.
  - g. Código en R que usa `summarise`.
  - h. Código en R que usa `group_by`.
  - i. Código en R que usa `filter`.
  - j. Código en R que usa `select`.
  - k. Escritura del reporte. *Esto refiere al reporte inicial (al borrador) puede ser que todos escriban su parte.*
  - l. Edición y redacción final del reporte. *Es la persona que se encarga de revisar que no tenga faltas de ortografía y que todo se lea de manera cohesiva. Es quien pasa de borrador a reporte final que se envía al profesor*
  - m. Organización del equipo. *Determinar quién se encarga de qué parte así como fechas de entrega.*
9. Código obtenido de una página o de un compañero debe citarse como comentario.
10. El reporte debe incluir al menos una cita bibliográfica referente al tema que trata (es decir si tu base es

de delincuencia incluir al menos una fuente que explique cuál es el problema de delincuencia en México ya sea libro, revista, artículo, página web). Todo el texto debe estar adecuadamente citado con la referencia (pueden elegir el formato MLA, APA, Chicago, da igual).

11. El proyecto debe presentarse el 27 ó 29 de septiembre por zoom durante clase en una duración de a lo más 3 minutos por proyecto (piensen en un video de TikTok).

## Evaluación

Hay dos componentes para la evaluación: individual y del equipo.

**Individual (50%)** - Las gráficas con el nombre del alumno cumplen con los puntos estilísticos especificados arriba (1 pto). - Las gráficas con el nombre del alumno son adecuadas para los datos y la explicación de lo que representa es correcta (2 ptos) - La parte realizada por el alumno de los incisos a a la j cumple con los requisitos de arriba (por ejemplo código siguiendo las normas de tidyverse) y las explicaciones involucradas tienen sentido (2 ptos)

**Grupal (50%)** - Todas las gráficas cumplen los puntos estilísticos (0.5 pto). - El texto explica claramente qué se está analizando y por qué es importante (0.5 pto). - Los estadísticos utilizados tienen sentido para los datos involucrados (2 ptos). - El código de **R** corre a la perfección, sigue las normas de tidyverse y además estima las cantidades correctas (1 pto) - La exposición del tema fue en tiempo y explica: qué querían analizar de los datos y qué encontraron en términos **muy generales** (NO ES NECESARIO PPT PUEDE SER TODO ORAL O CON SU PDF)

**NOTA** A partir de los 5 errores ortográficos comienzo a restar un punto (general) por cada error.