

Informe Encuesta Metodología y Técnicas Cuantitativas

Metodología Cuantitativa

15/2/23

Tabla de contenidos

Presentación y objetivos del informe

El siguiente informe tiene una finalidad principalmente pedagógica. Específicamente está pensado como un mecanismo para ayudar a alcanzar los siguientes objetivos que forman parte del [programa](#) de la materia “Metodología y Técnicas Cuantitativas” de la UNAJ:

- Lograr un conocimiento mínimo de la existencia y pertinencia de las técnicas de análisis de datos *básicas* y una *habilidad* mínima en la ejecución de las mismas.
- *Lograr un* conocimiento mínimo de la existencia y pertinencia de (otras) técnicas de análisis de datos más específicas y (usualmente) más *complejas*.

Aparte de la distinción entre técnicas básicas y complejas la diferencia en cuanto a los objetivos pedagógicos es que entre ambos objetivos es que las últimas sólo se aspira a conocerlas mientras que en las primeras se espera, también, que se adquiriera cierta habilidad en su ejecución. Expresado de otro modo, de las técnicas consideradas complejas se espera que se sepa de su existencia y qué tipos de problemas ayuda a solucionar. En cambio, en las técnicas básicas se espera lo anterior pero que además se adquiriera la habilidad de poder ejecutar las mismas.

Para lograr lo anterior durante la cursada de la materia se realiza una encuesta mediante un [formulario de google](#) que contestan los mismos estudiantes. Luego el contenido de sus respuestas se analiza con el [lenguaje R](#) y con el mismo se escribe y publica este informe.

En cuanto al estilo del texto será deliberadamente informal aunque con 3 excepciones con fines principalmente pedagógicos:

- Se incluirán piezas de código
- Se hará uso de citas y referencias (estilo APA)
- Los conceptos importantes aparecerán traducidos al inglés entre paréntesis.

Por otro lado, el texto estará acompañado de los siguientes elementos visuales:

1 Diseño de la encuesta

En el diseño de la encuesta que contestan los estudiantes se pueden distinguir 2 grandes procesos. Uno referido al diseño del cuestionario y otro referido al diseño de la selección de los casos. El primero se preocupa por **qué se pregunta** y el segundo por **a quienes**. En esta sección la mayoría de las veces sólo se explicitará pero pocas veces se justificará las decisiones tomadas en ambos procesos. En otras palabras, la palabra “diseño” le queda grande a ambos procesos para esta encuesta. Lamentablemente lo mismo puede afirmarse de varias encuestas académicas o profesionales. La justificación particular para la falta de atención metodológica a estos puntos es que en este caso se trata de una actividad con fines principalmente pedagógicos más que académicos o profesionales.

1.1 Diseño del cuestionario

El [cuestionario](#), como muchos cuestionarios que se usan en la práctica profesional, está diseñado con la lógica de módulos. Como veremos más adelante en esta instancia cada módulo tiene una justificable lógica interna pero lo que no tiene este cuestionario es una coherencia interna que explique la relación entre los diferentes módulos.

El diseño de un cuestionario mediante la estrategia de módulos suele ser útil tanto para el encuestado como para el investigador. Al encuestado lo ayuda a ordenarse visualmente especialmente si se trata de un cuestionario largo. Al investigador lo ayuda tanto para el etiquetamiento (o nombramiento) como para el orden de las variables en la base de datos.

El cuestionario por ahora cuenta con los siguientes módulos:

1. Identificación
2. Demográfico
3. Composición del hogar
4. Cuidados
5. Vivienda
6. Uso del tiempo
7. Inseguridad alimentaria

8. Preferencias sociales
9. Origen social
10. Trabajo actual
11. Ingresos del hogar
12. Académico UNAJ
13. Expectativas materia cuantitativa
14. Redes sociales entre estudiantes de la materia

Claramente no parece haber mucha relación entre los diferentes módulos. A diferencia de algunas encuestas ómnibus en donde este problema se presenta porque cada investigador que forma parte de la investigación agrega su propio “paquete de preguntas” para su particular investigación en este caso la justificación es que cada módulo tiene una función pedagógica particular.¹

1.2 Diseño de la selección de casos

En cuanto al criterio utilizado para la selección de los casos se puede afirmar que se trata de una muestra de los estudiantes de la materia “Metodología y Técnicas Cuantitativas” de la carrera de Trabajo Social de la Universidad Nacional Arturo Jauretche. Esto ya afirma algo pero se puede especificar aún más.

Antes que nada, si bien es una encuesta que potencialmente les llega a todos los estudiantes que cursan la materia, la misma efectivamente sólo llega a una muestra de los mismos. Esto se produce principalmente por la no-respuesta de algunos estudiantes que comienzan la cursada pero no responden la encuesta. El factor “deserción” no parece afectar tanto a la selección de casos dado que la encuesta se realiza al principio de la cursada. Esto hace que la encuesta, si es representativa de algo, lo sea de la población intermedia que queda conformada entre los inscriptos a la materia y los que finalmente la regularizan. Obviamente esto no alcanza para decir mucho acerca de la representatividad de la encuesta sobre poblaciones mayores compuestas por:

- Los estudiantes de la carrera de Trabajo Social de la UNAJ.
- Los estudiantes de la UNAJ
- Los estudiantes del sistema universitario nacional

¹En una investigación ómnibus varios investigadores (o clientes) comparten el mismo diseño de selección de casos y de esa manera, juntos, pueden llegar a más casos ya que entre todos amortizan estos costos que suelen ser grandes en encuestas grandes presenciales. A cambio, más allá de un algunos módulos comunes (p.e. el demográfico), cada investigador o cliente agrega su módulo de particular interés.

2 Limpieza y consistencia de los datos

El producto de los procesos de producción y registro de los datos suele cristalizarse en una (o varias) bases de datos. Estas pueden (y suelen) contener diferentes tipos de errores por lo que se considera una buena práctica realizar un proceso de limpieza y preparación para recién después comenzar el proceso estricto del análisis de los mismas.

En esta sección veremos algunos ejemplos tanto de limpieza, consistencia y construcción de nuevas variables. Aquí veremos ejemplos de los casos mas sencillos. Procesos como el pegado (*joint*) de variables, necesario cuando los datos se encuentran en diferentes archivos o bases de datos, no se verán aquí.

2.1 Limpieza

La idea de limpieza (*cleaning*) viene de usar la metáfora de dato sucio (*dirty*). Un dato sucio no necesariamente es un dato incorrecto aunque sí se trata de un tipo de dato más incómodo de trabajar ya que dificulta el posterior análisis.

La tarea de la limpieza de una base de datos suele implicar un tiempo, especialmente en bases de datos con muchas variables (muchas columnas). Si bien es una práctica recomendada siempre hay que tener en cuenta la “escala” del trabajo a realizar porque que en algunas situaciones puede que sea más simple corregir algunas cuestiones de manera artesanal o a mano en el proceso mismo de la publicación de los análisis. Esto suele ser particularmente cierto en algunos de los siguientes escenarios y sus posibles combinaciones:

- Encuestas con muchas preguntas en donde se sabe que se van a analizar una sola vez o que se analiza sólo una pequeña parte de la información disponible.
- Base de datos y procesos de análisis que luego no se van a compartir (los resultados se comparten pero estos no son replicables por terceros)

De manera complementaria el proceso de limpieza se amortiza considerablemente cuando los datos se van a analizar más de una vez y cuando se requiere un grado de transparencia que exige que la investigación sea enteramente replicable.

2.1.1 Renombre de las variables

La tarea básica de limpieza (*cleaning*) que aquí se hará será el **renombre de todas las variables**. La razón de esta operación es que, al menos si se trabaja con google forms (aunque algo similar suele suceder con otros sistemas de formularios online), los nombres de las variables son el texto de la propia pregunta del formulario. Esto incomoda un poco el análisis de los datos por la gran extensión de algunas preguntas y esa incomodidad se traduce en una mayor propensión al error. En este sentido para trabajar con los nombres de las variables suele ser recomendable:

- Eliminar los espacios entre las palabras agregando algún símbolo que las pegue o una (“_”, “-” o cualquier otro),
- Pasar todas las letras a minúscula (o mayúscula)
- Renombrar el nombre original con un nombre más corto y recordable. Una opción recomendable es que cuando en la encuesta haya variables que corresponde a un mismo módulo (p.e. módulo vivienda) se inicien por un mismo prefijo (viv_habitaciones, viv_inodoro, etc.).

2.1.2 Orden de las variables

Otra tarea diferente pero relacionada con la anterior tiene que ver con el **orden de las variables** en la base de datos. En algunos sistemas el orden de las variables tiene que ver con el orden cronológico en que se fueron construyendo las preguntas (p.e. en una encuesta) o con el orden en que luego se fueron construyendo variables complejas o recategorizaciones. En cualquier caso, lo que se debe tratar de lograr es que las variables que tienen relación temática entre sí, no sólo tengan un prefijo que las una sino que se encuentren visualmente cercanas en la base de datos. Esta recomendación es más importante cuanto más variables tenga la base de datos.

2.1.3 Etiquetado de las variables

Antes se habló de renombrar y ordenar las variables. Sin embargo, en las ciencias sociales y en especial en aquellas disciplinas donde se encuentren difundidos programas estadísticos como el SPSS, SAS y Stata se usa la distinción entre **nombre de la variable y etiqueta de la misma**. El primero es como el nombre real de la variable y así lo entiende el mismo programa. La segunda es como un alias o un metadato que permite una interpretación más humana del significado de la variable. Muchas veces el contenido de la etiqueta se acerca a la pregunta original. Ahora bien, las variables, especialmente las que se suelen denominar discretas o categóricas, aparte de tener un nombre pueden contener una serie finita de categorías.

Cuando se trata de variables del tipo “Indique su cantidad de hijos” si la respuesta de la base de datos es un “2” se entiende que la persona ha respondido que tiene 2 hijos. Lo mismo si aparece un “1” o un “3”. Se dice que estos son los **valores** de las variables. Pero el problema comienza, siguiendo una vieja tradición del análisis de datos (ver “De donde vienen las etiquetas”) cuando como respuesta a variables del tipo “Indique su género” nos encontramos con símbolos (más precisamente numerales) como “1”, “2”, etc. en la base de datos. Estos símbolos se suelen llamar **códigos** y, sin información externa, no hay manera de saber que significan. Para eso vuelven en nuestra ayuda las etiquetas

Así como hay etiquetas para los nombres de las variables también lo hay para sus etiquetas.

La etiqueta es especialmente útil en los contextos de presentación de los análisis sea tanto en formato de tablas y gráficos. La razón es que el lector de los informes puede no estar al tanto ni del cuestionario original ni de la propia base de datos. Simplemente es un lector de un informe que desea, razonablemente, que en vez de algo como “n_hijos” aparezca algo como “Cantidad de hijos” o que en vez de “3” en los análisis de la variable “Género” aparezca algo como “género no binario”.

2.1.4 Tipo de dato de la variable

Otro punto a destacar es lo que a veces se suele denominar “nivel de medición” de la variable. Este término se suele utilizar más en programas o lenguajes específicos para análisis de datos pero tienen una similitud con el proceso usual en una planilla de cálculo (excel, google sheet, etc) de aplicarle “formato” a la celda para indicarle si se trata de texto o un número y si se trata de número de que tipo (porcentaje, fecha, decimal, etc.).

Este tipo de metainformación suele ser útil para que el programa detecte que tipo de gráfico es apropiado o que tipo de cálculo realizar. De todos modos los programas vienen cada vez con mejores heurísticas para adivinar, sin previa indicación de que tipo de dato se trata.

3 Análisis Espaciales

La cartografía, la disciplina que se encarga de la confección y uso de mapas, es una actividad milenaria. Sin embargo, desde hace relativamente pocas décadas se ha comenzado a difundir el uso de técnicas que, no sólo representan a diferentes objetos en un mapa, sino que también, al mismo tiempo, permiten la realización de diferentes análisis específicamente espaciales.

Los mapas pueden considerarse como una representación gráfica basada en una proyección, que “proyecta” sobre un plano de 2 dimensiones (2D) una realidad tridimensional (3D). Esta representación permite hacer visible relaciones de distancia entre elementos en algún espacio. En particular, los mapas geográficos (o cartográficos) suponen unas propiedades métricas que dependen de la proyección utilizada, y posibilitan la toma de medidas de distancias, ángulos o superficies sobre él y su posterior (re)proyección a la realidad.

En la antigüedad, cuando sólo se sabía la ubicación específica de pocos elementos (no había GPS), se confiaba en la representación gráfica realizada sobre un plano (mapa) construido sobre esas (pocas) ubicaciones conocidas y se calculaba la posición de los elementos todavía no posicionados, sea tanto una nueva ciudad conocida o la propia ubicación de un viajante.

i Los mapas en la historia

En la actualidad, gracias a la tecnología satelital, se puede tener información de la ubicación específica de (casi) todo lo que a uno le interese y . De todos modos, la parte conceptual que se encuentra por detrás de lo anterior es sumamente antigua y En efecto, su creador fue nuestro viejo amigo [Eratóstenes](#). Él ideó lo que se considera el primer sistema de coordenadas geográfica en donde, bajo el supuesto de la esferidad perfecta (un cuerpo), con sólo 2 datos se podría determinar cualquier posición.

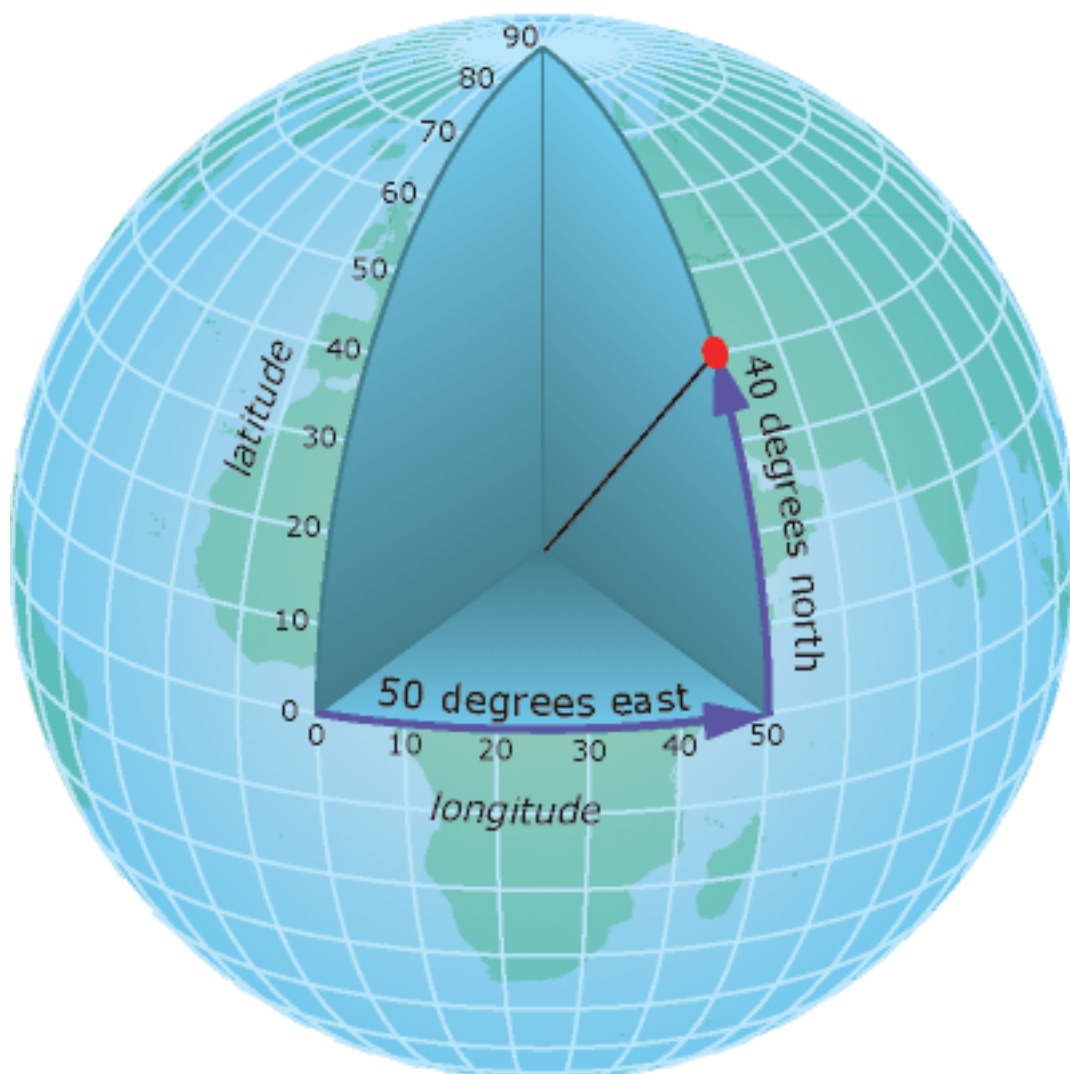


Figura 3.1: Sistemas de coordenadas teniendo como referencia una esfera o algún tipo de elipsoide.

De modo conceptualmente independiente, para localizar esa posición en un mapa de 2 dimensiones (plano) se requiere una proyección, esto es, una relación ordenada entre los puntos de la superficie curva de la Tierra, (o en general de cualquier cuerpo) y los de un mapa, (o en general de cualquier plano).

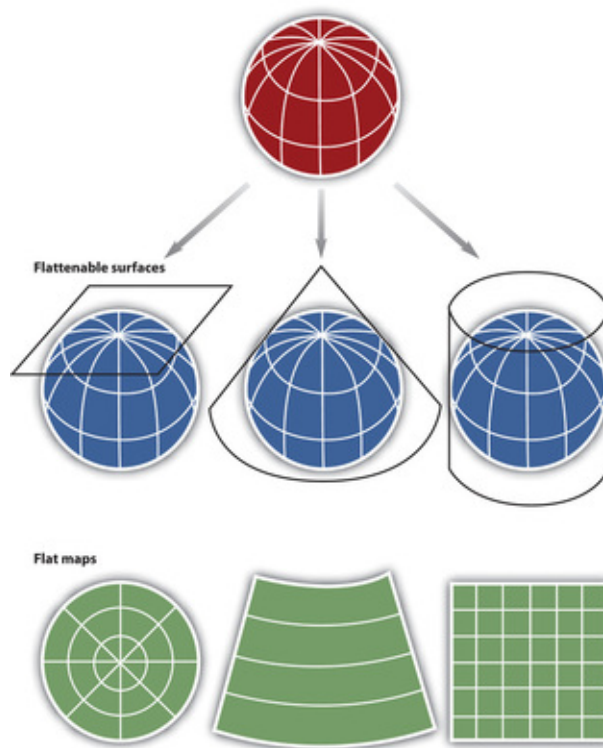


Figura 3.2: Proyecciones genéricas. La 3ra es la más similar a la usual proyección de Mercator

Es este sentido, es importante aclarar que todo mapa geográfico de 2 dimensiones implica alguna proyección y que, dependiendo cual sea está última, genera un tipo de distorsión específica. La distorsión se genera al pasar de un mundo con 3 dimensiones (cuerpo geométrico) a un plano con 2 dimensiones (figura geométrica). Por ejemplo la [proyección Mercator](#), la usual cuando un@ va (¿o iba?) a una librería a comprar un mapa n° 3, supone que Groenlandia aparece aproximadamente del tamaño de África, cuando en realidad el área de África es aproximadamente 14 veces la de Groenlandia. Lo anterior no sucede en los globos terráqueos ya que estos no son proyecciones.



Figura 3.3: El erdapfel es el globo terráqueo más antiguo que aún se conserva. Lo construyó un cartógrafo alemán llamado Martin Behaim en 1492

Aún con sus imperfecciones, los mapas planos también son de gran utilidad por fuera de la navegación. En efecto, pueden servir para hacer observables hechos que antes no lo eran, así como para el diseño, ejecución y evaluación de políticas territoriales. Por ejemplo, [John Snow](#), uno de los fundadores de la epidemiología, logró trazar el origen del cólera siguiendo la pista de los enfermos a través de un mapa. Con esta ayuda hipotetizó que el agua en mal estado era un vector importante en la propagación de la enfermedad cuando por aquel entonces se suponía que el cólera (al igual que la peste bubónica) tenía que ver con la polución del aire más que con el agua.

Snow llegó a su conclusión luego de «georeferenciar» los fallecimientos (gracias a que un cura amigo averiguó donde vivía esa gente!) y observar la cercanía de la mayor cantidad de muertes con una bomba manual de agua pública. Recomendó retirar la manija de la bomba y a los pocos días la epidemia cesó. Posteriormente se supo que los desechos de

uno de los pozos ciegos de las casas linderas se había filtrado hacia la capa freática que la bomba extraía el agua.



Figura 3.4: Mapa original de John Snow (1854). Los puntos negros simbolizan fallecimientos.

3.1 Insumos de la encuesta

Los principales análisis que se muestran a continuación se basan en una serie de preguntas del módulo de vivienda que refieren a la dirección de la misma como la calle, la altura y el partido

o departamento del estudiante.

Sección 11 de 23

Módulo Vivienda

A continuación se le harán una serie de preguntas referidas a la vivienda en donde usualmente habita usted y su hogar.

Con respecto a la ubicación de la misma ¿Nos podrias indicar los siguientes datos...?

Descripción (opcional)

la calle de tu vivienda es... *

Texto de respuesta breve

el número de la altura de la calle es... *

Texto de respuesta breve

esa vivienda se encuentra en el partido de... *

☐ Almirante Brown

☐ Avellaneda

☐ Berazategui

Figura 3.5: Preguntas sobre ubicación de la vivienda

3.2 Análisis espaciales

Todos solemos tener intuiciones acerca de como el espacio influye en las decisiones de las personas.

La georefeenciación es un proceso que consta en asignarle unas coordenadas geográficas específicas a algún hecho, aun cuando este cambie en tiempo y espacio, como por ejemplo, la trayectoria recorrida en una bicicleta. Todo lo que se pueda decir que sucede en un espacio geográfico es plausible de aplicarle una georeferenciación. La cantidad de información georeferenciada viene creciendo a pasos agigantados en los últimos tiempos. Esto es especialmente notoria en ciertos datos importantes que no cambian seguido en el tiempo, como por ejemplo, las divisiones políticas de los países. Este tipo de datos ya suelen ser de libre disponibilidad y son un soporte genérico para cualquier investigación específica.

El talón de Aquiles de una investigación que quiera hacer uso de los mapas modernos es que cada investigador debe agregar los datos específicos de su investigación que quiera georeferenciar. Expresado en la jerga actual, se dice que cada investigador debe agregar una *capa* específica que contenga sus datos para luego, si lo desea, relacionarla con los otros datos genéricos que se almacenan en otras capas genéricas. Para fijar las ideas, si yo quiero realizar una investigación con los estudiantes de la Universidad voy a tener que tener georeferenciada la ubicación de cada uno de ellos. Como sucede con cualquier observación o medición, cualquier dato preciso se puede agrupar después pero la inversa no es posible. Por ejemplo si tengo el municipio en donde vive cada estudiantes podré hacer un mapa que coloree los polígonos de los municipios en función de la cantidad de estudiantes (*Colorpleth*) pero no podré hacer un [mapa de puntos](#), porque este requiere las coordenadas exactas donde vive cada estudiante. En cambio, si tengo las coordenadas puedo hacer tanto el mapa de puntos como una graficación de los polígonos de cada municipio.

La tecnología que permite manejar y dibujar un mapa trabajando con varias capas de información georeferenciada se llama GIS (Geographical Information System). Históricamente De forma más contemporánea, los GIS online, también permitieron que no sea (tan) necesario como antes la construcción de mapas sobre un plano mediante una única proyección. La razón es que en la actualidad, la conocida función de «zoom» de programas como google maps (o su antecesor el google earth) ajusta de modo automático entre diferentes proyecciones. En otras palabras, se actualiza una proyección diferente para cada nivel de zoom.

En otras palabras, los mapas online de ahora son la unión de:

- Una mejor información satelital en cuanto a fotos satelitales y cálculos de distancias.
- Una gran cantidad de información georeferenciada (capas), muchas (ahora) construidas por voluntarios desde sus teléfonos celulares
- Una mejor procesamiento digital de mayor información disponible, mediante programas GIS y (ahora) también web, que vuelve menos relevante la construcción de un mapa sobre un plano.

4 Análisis de textos

4.1 Insumos de la encuesta

El insumo de estos análisis es una serie de preguntas de respuesta abierta. Sin embargo, dado el formato específico de la pregunta se trata de un dato que, si bien no puede considerarse estructurado, se aleja un poco de otros ejemplos de datos no estructurados como los textos provenientes de entrevistas en profundidad, de canciones o de discursos presidenciales.

The image shows a digital survey form interface. At the top, there is a question box with a light blue border and a vertical blue bar on the left. The question text is: "A continuación nos interesa que nos digas 3 palabras que se te vienen a al cabeza cuando vos escuchas la frase "Metodología y Técnicas Cuantitativas". No importa si son palabras que vos consideras como positivas o como negativas." Below the question box is a label "Descripción (opcional)" followed by a horizontal line. To the right of the question box is a vertical toolbar with icons for adding, copying, deleting, and other actions. Below the question box are three input fields, each with a label and a red asterisk: "Primer palabra *", "Segunda palabra *", and "Tercera palabra *". Each input field has a placeholder text "Texto de respuesta breve" and a horizontal line for text entry.

Figura 4.1: pregunta_texto

En este caso, como puede observarse en la imagen se trata de 3 palabras por cada encuestado. Expresado en el léxico de una matriz de datos (o base de casos x variables), se trata de 3 columnas por cada fila.

4.2 Preparación de los datos

Para facilitar el siguiente análisis lo que primero debemos realizar es un “alargamiento” de la base de datos original para pasar a tener una sola columna y más filas que, en principio, se repetirían 3 veces.

Luego, sigue el proceso de limpieza de las palabras. Se pasa todo a minúscula, se eliminan los puntos, los números y toda palabra vacía de significado como los pronombres, artículos y preposiciones. Finalmente se hace un proceso en donde se intenta llevar todas las palabras a su raíz (*stemming*) para así poder interpretar como una misma palabra a palabras que, por ejemplo, cambian el género (gato, gata) o el número (gato, gatos). Ahora sí, ya estamos en mejores condiciones de empezar nuestro análisis del texto.

4.3 Primeras aproximaciones

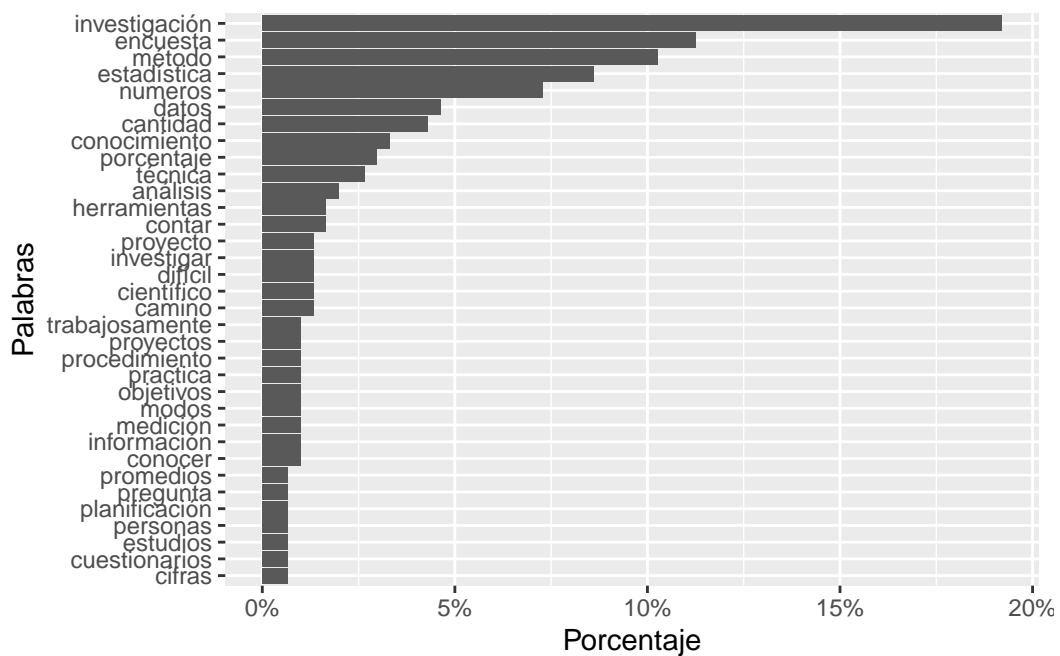
Dado que los pasos anteriores nos permitieron estructurar bastante los datos, ahora es más fácil realizar los típicos análisis que se realizan a los datos estructurados. Para comenzar vamos a realizar una simple tabla de conteo, con su respectivo porcentaje, y luego vamos a realizar un gráfico de barras. Finalmente realizaremos una nube de palabras.

Frecuencia y porcentajes de palabras

Palabra	Cantidad	Porcentaje
investigación	58	19.21%
encuesta	34	11.26%
método	31	10.26%
estadística	26	8.61%
numeros	22	7.28%
datos	14	4.64%
cantidad	13	4.30%
conocimiento	10	3.31%
porcentaje	9	2.98%
técnica	8	2.65%
análisis	6	1.99%
contar	5	1.66%
herramientas	5	1.66%
camino	4	1.32%
científico	4	1.32%
difícil	4	1.32%
investigar	4	1.32%
proyecto	4	1.32%

conocer	3	0.99%
información	3	0.99%
medición	3	0.99%
modos	3	0.99%
objetivos	3	0.99%
practica	3	0.99%
procedimiento	3	0.99%
proyectos	3	0.99%
trabajosamente	3	0.99%
cifras	2	0.66%
cuestionarios	2	0.66%
estudios	2	0.66%
personas	2	0.66%
planificación	2	0.66%
pregunta	2	0.66%
promedios	2	0.66%

Como en muchas otras situaciones, lo que se puede mostrar en forma de tabla también se puede graficar con algún tipo de gráfico. En este caso haremos un gráfico de barras con los datos anteriores.



4.4 Nube de palabras

La nube de palabras es una técnica de visualización que funciona bien cuando los insumos son palabras y están presentando una gran heterogeneidad en los valores de sus frecuencias.

Nube de palabras



4.5 Análisis bivariado de palabras

Frecuencia y porcentajes de palabras según cantidad de materias aprobadas

	Cantidad de materias aprobadas	
Palabra	Hasta 15	Más de 15
investigación	20.00%	20.00%
encuesta	12.22%	10.48%
método	11.11%	10.48%
estadística	7.78%	11.43%
numeros	7.78%	6.67%
cantidad	4.44%	4.76%
datos	3.33%	7.62%
técnica	3.89%	NA
conocimiento	2.78%	4.76%

porcentaje	2.78%	3.81%
contar	2.22%	NA
investigar	2.22%	NA
científico	NA	3.81%
análisis	1.67%	2.86%
camino	1.67%	NA
conocer	1.67%	NA
difícil	1.67%	NA
proyectos	1.67%	NA
herramientas	1.11%	2.86%
trabajosamente	NA	2.86%
información	1.11%	NA
medición	1.11%	NA
modos	1.11%	NA
personas	1.11%	NA
planificación	1.11%	NA
pregunta	1.11%	NA
procedimiento	1.11%	NA
promedios	1.11%	NA
proyecto	1.11%	1.90%
estudios	NA	1.90%
objetivos	NA	1.90%
practica	NA	1.90%

Nube de palabra según cantidad de materias aprobadas



5 Análisis de redes

5.1 Insumos de la encuesta

El insumo de estos análisis es una serie de preguntas en donde cada estudiante debe elegir, entre todos los estudiantes de ese cuatrimestre, a los 5 a los que más conoce. Como tal este módulo puede considerarse como una serie de preguntas cerradas de respuesta única con la característica saliente de que esas preguntas poseen muchas opciones posibles. En efecto, las opciones son tantas como estudiantes estén cursando la materia. Desde un punto de vista metodológico lo importante no es tanto que las opciones sean muchas o pocas sino que el rango de estas es conocido y discreto.

Luego de una primera pregunta en donde el estudiante se auto identifica en la lista se pasa a otras 5 preguntas en donde el estudiante va eligiendo a otros compañeros a los cuales más conoce. Si no conoce a nadie más elige la opción “00-NO CONOZCO A NADIE MAS” en las siguientes preguntas.

La explicitación de lo anterior es importante por la siguiente razón. El análisis de redes sociales (SNA) no se suele llevar bien con la estructura de datos conocida conocida en las ciencias sociales como **matriz de datos** (Galtung 1973) o, más en general, como una estructura de datos “casos x variables”. Como lo sugiere esta última descripción esta es una estructura afín a lo que se suele denominar “datos de atributos” en donde los valores de las variables se consideran atributos de cada uno de los casos. En cambio, en el SNA suele ser preferible alguna de las siguientes estructura de datos que se consideran más idóneas para representar un tipo de dato **relacional** (Scott 2017, cap. 4):

a) Una **matriz de adyacencia** (*adjacency matrix*) en donde las filas y las columnas son asignados para los (mismos) nodos de la red y la presencia de una relación es representada con un valor numérico. Se dice que es una matriz cuadrada y dispersa porque tiene la misma cantidad de filas y columnas y puede que muchas de sus celdas contengan un valor “0”. Se suele considerar la representación estándar de esta técnica.

b) Una **lista de lazos** (*edge list*) en donde (sólo) se “listan” todas las relaciones de la red. Es una opción que, en comparación a la opción “a”, suele ocupar menos espacio en memoria ya que sólo representa la presencia y/o intensidad de las relaciones entre los nodos. En general se trata de una estructura más larga que ancha ya que cada relación ocupa una fila y las columnas se suelen limitar (aunque no necesariamente) a algo como “origen” (*from*), “destino” (*to*) e “intensidad de la relación”.

Sección 23 de 23

Módulo Análisis de Redes Sociales

A continuación te vamos a pedir algunos datos para construir un gráfico de las relaciones sociales entre los estudiantes de la materia. Para eso te mostraremos la lista de los estudiantes (ordenados alfabéticamente por su nombre) que se han anotado en este cuatrimestre para cursar Metodología y Técnicas Cuantitativas.

En la primer pregunta debes seleccionarte a vos mismo en la lista.

Luego debes comenzar a elegir a los 5 estudiantes que más conozcas. Para eso tendrás que responder otras 5 preguntas y seleccionar en la primera de ellas al/a que más conoces de esos 5 y luego, en orden decreciente, seguir completando las preguntas para cada uno de los otros 4 estudiantes que más conozcas. Si no conoces a nadie podes elegir la opción "00 NO CONOZCO A NADIE MAS".

En la lista habrá estudiantes que no pertenecen a tu comisión pero que sí están cursando este cuatrimestre la materia. Si los consideras entre tus 5 contactos más cercanos, los podes elegir.

Por favor, selecciona TU NOMBRE dentro de esta lista *

1. ALDANA, [REDACTED]
2. ANDREA JOHANA, [REDACTED]
3. ANTONELLA SOLEDAD, [REDACTED]

Figura 5.1: pregunta_redes_1

Selecciona al PRIMER estudiante *

1. 00 - NO CONOZCO A NADIE MAS
2. ALDANA, [REDACTED]
3. ANDREA JOHANA, [REDACTED]
4. ANTONELLA SOLEDAD, [REDACTED]

Figura 5.2: pregunta_redes_2