



Islas Canarias
Del 15 al 19 de noviembre de 2021



Escuela Andaluza
de Salud Pública
Consejería de Salud y Familias



UNIVERSIDAD
DE GRANADA



XXI Jornadas de Estadística de las Comunidades Autónomas

Encuesta Sanitaria y Social de Hogares (ESSOC): análisis y visualización mediante R y Python

Luis Castro Martín
María del Mar Rueda García
Andrés Cabrera León
Carmen Sánchez-Cantalejo Garrido
Ramón Ferri García
Jorge Hidalgo Calderón

Introducción

Necesidad de comprender el impacto del COVID-19  ESSOC.

ESSOC recoge información sobre características sanitarias, socioeconómicas, psicosociales, conductuales, laborales, medioambientales y clínicas de la población general y de colectivos vulnerables.

Datos provenientes de encuestas y registros clínicos. Encuesta tipo panel por superposición.

Introducción

Se puede observar evolución de ciertas características, aunque existen problemas de no respuesta.

El problema de la no respuesta se corrige utilizando técnicas de reponderación usando información auxiliar disponible.

Después de los ajustes en los pesos muestrales para reducir posibles sesgos debidos a la falta de respuesta, se generan estimaciones que se almacenan en 6 tablas descriptivas.

Una gran cantidad de información que requiere de su visualización en gráficos para posibilitar su interpretación.

Objetivos

Mostrar las distintas **tablas descriptivas** que resumen la información de la ESSOC.

Motivar y explicar el uso de la librería **Altair** de Python para visualizar gráficamente la información de las tablas descriptivas de la ESSOC.

Describir el desarrollo de la **Web de la ESSOC** donde se almacenarán las tablas descriptivas y sus correspondientes gráficas.

Resultados: Tablas descriptivas

Problema: Más de 1000 variables que generan miles de tablas con decenas de resultados en su interior cada una.

Pregunta: ¿Qué podemos hacer para digerirlos, identificar la información relevante y poder así plantear hipótesis y/o conclusiones?

Respuesta: visualización de datos.

Transversal Variables Originales

Autopercepción de salud general		Total					Hombres				
		n (muestra)	N (estimada)	Estimación puntual ()	lim inf IC	lim sup IC	n (muestra)	N (estimada)	Estimación puntual ()	lim inf IC	lim sup IC
Todas las edades	Excelente	513	1211266	17.16	15.72	18.61	283	710780	20.61	18.28	22.93
	Muy buena	579	1349099	19.12	17.62	20.62	271	679225	19.69	17.39	21.99
	Buena	1391	3302206	46.8	44.86	48.73	639	1618496	46.92	44	49.84
	Regular	403	1034481	14.66	13.23	16.09	135	385679	11.18	9.18	13.18
	Mala	62	159624	2.26	1.68	2.84	21	55319	1.6	0.9	2.3
	Total	2948	7056676	100	*	*	1349	3449501	100	*	*
16-24	Excelente	122	285429	35.34	30.06	40.61	69	162069	38.91	31.27	46.56
	Muy buena	84	192475	23.83	19.19	28.46	38	88045	21.14	14.96	27.32
	Buena	111	276227	34.2	28.64	39.75	57	152306	36.57	28.35	44.79
	Regular	21	51095	6.33	3.62	9.03	6	14053	3.37	0.69	6.06
	Mala	1	2538	0.31	-0.3	0.93	0	0	0	0	0
	Total	339	807767	100	*	*	170	416475	100	*	*
25-34	Excelente	106	260233	25.87	21.38	30.36	54	141438	27.7	20.89	34.51
	Muy buena	105	266468	26.49	21.62	31.36	46	135801	26.59	18.96	34.23
	Buena	152	403168	40.08	34.65	45.5	69	198741	38.92	30.61	47.23
	Regular	28	76121	7.57	4.61	10.52	11	34701	6.8	2.4	11.19
	Mala	0	0	0	0	0	0	0	0	0	0
	Total	391	1005991	100	*	*	180	510684	100	*	*

Transversal Variables Dicotomizadas

Sentimiento de soledad poco frecuente (1 o 2)	Total				Hombres				Mujeres			
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)	Estimación porcentual (IC95)
Total	20.41 (18.92-21.99)	19.24 (17.54-21.07)	23.36 (21.76-25.05)	25.86 (24.1-27.71)	16.42 (14.44-18.61)	15.95 (13.47-18.79)	19.61 (17.34-22.09)	20.8 (18.18-23.69)	24.23 (22.08-26.51)	22.38 (20.12-24.82)	26.94 (24.73-29.27)	30.69 (28.37-33.12)
Urbano	22.09 (19.94-24.4)	20.54 (18.06-23.26)	23.96 (21.73-26.34)	26.03 (23.52-28.71)	16.02 (12.96-19.65)	14.18 (10.64-18.64)	19.55 (15.97-23.71)	21.25 (17.25-25.88)	22.86 (19.57-26.51)	22.81 (19.2-26.88)	27.04 (23.5-30.91)	32.02 (28.25-36.05)
Intermedio	19.43 (17.12-21.96)	18.52 (15.92-21.45)	23.29 (20.72-26.08)	26.64 (23.83-29.64)	15.1 (10.01-22.15)	12.91 (7.79-20.64)	19.93 (13.5-28.43)	16.2 (11.03-23.17)	18.33 (12.92-25.34)	19.66 (13.44-27.85)	22.24 (16.28-29.59)	29.06 (22.15-37.1)
Rural	16.69 (12.8-21.47)	16.23 (11.98-21.62)	21.08 (16.53-26.48)	22.57 (18.04-27.85)	17.09 (14.31-20.29)	18.2 (14.43-22.68)	19.57 (16.49-23.07)	21.6 (17.71-26.09)	26.57 (23.48-29.92)	22.67 (19.56-26.11)	27.91 (24.84-31.21)	30.06 (26.92-33.4)

Resto de tablas:

Transversal Cambio y Transversal Cambio Brecha de Género.

Longitudinal Diferencia y Longitudinal Diferencia Brecha de Género.

Librería Altair (Python)

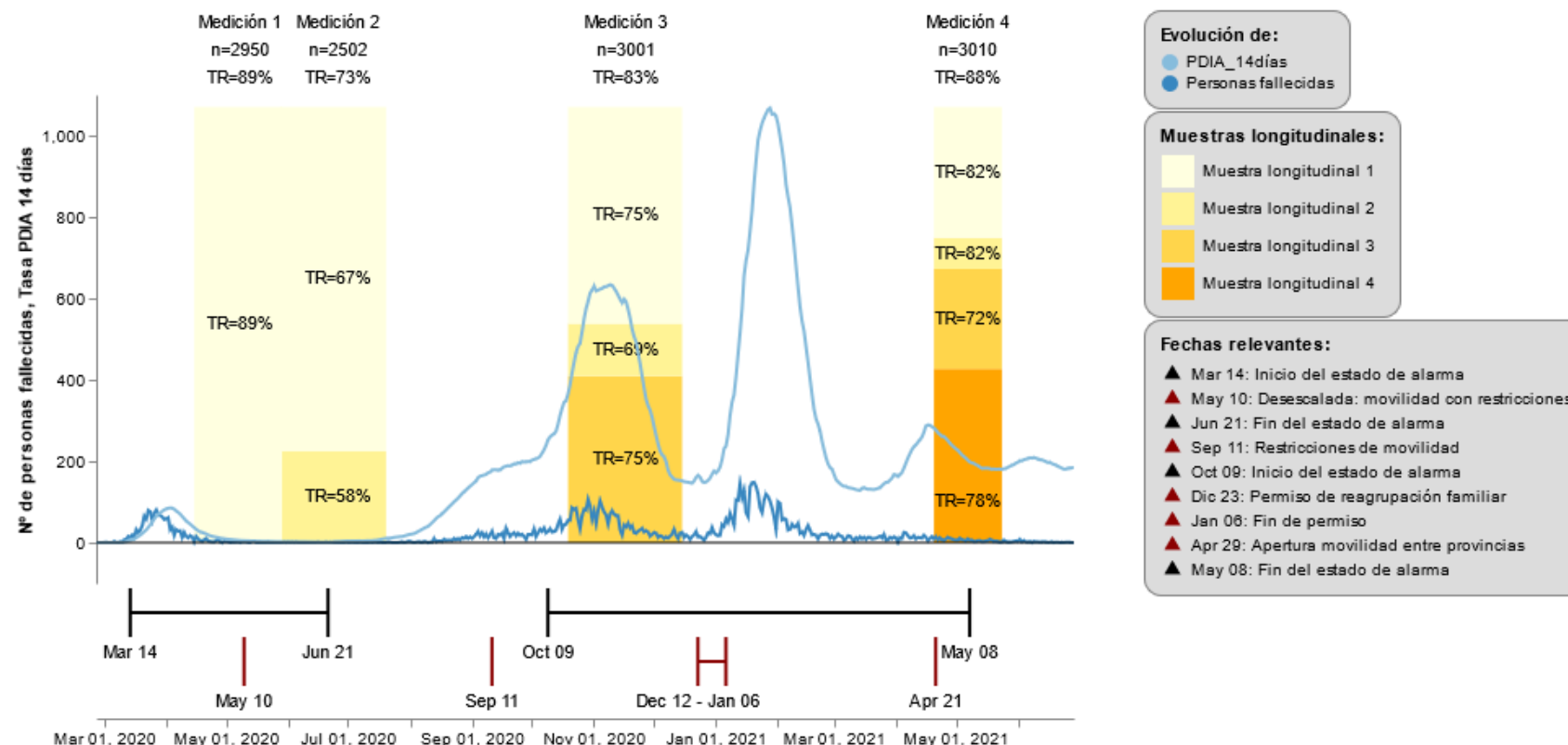
Encuesta Sanitaria y Social: trabajo de campo y evolución de la pandemia por SARS-COV-2 en Andalucía

n: Muestra efectiva, TR: Tasa de respuesta, PDIA: Pruebas Diagnósticas de Infección Activa

Fuente: IECA (Instituto de Estadística y Cartografía de Andalucía)

Ventajas:

- Personalización
- Interactividad



Visualización de datos

Tablas => Gráficas (NO)

Tablas => Procesamiento => Gráficas (SÍ)

¿Por qué es necesario un procesamiento? Aparte de una pequeña limpieza de las tablas, producidas con R, hay que solventar:

Problema 1: Trabajar con todas las variables de estratificación a la vez.

Problema 2: Variables de estratificación cruzadas en las tablas. Además la medición no es una variable.

Solución del problema 1

Crear un objeto donde se puedan referenciar todas las distintas tablas a la vez para trabajar conjuntamente. La lista:

$$df[l][i] \text{ para } l \in [0,1,2], i \in [0,1],$$

es estimación porcentual para $i = 0$, total para $i = 1$, y la variable Edad para $l = 0$, Provincia para $l = 1$, y Grado de Urbanización para $l = 2$.

Solución del problema 2

Desagregar las variables de estratificación y hacer de la medición una variable.

Se usa el atributo `.melt` de dataframes en pandas. Se pasa de esto:

In [20]: `df[0][0]`

Out [20]:

	autop	total_M2- M1_Porc	total_M3- M2_Porc	total_M4- M3_Porc	hombres_M2- M1_Porc	hombres_M3- M2_Porc	hombres_M4- M3_Porc	mujeres_M2- M1_Porc	mujeres_M3- M2_Porc	mujeres_M4- M3_Porc	edad
0	mejora la salud	0.209265	0.208286	0.208286	0.194489	0.238484	0.238484	0.223426	0.17935	0.17935	Total
1	NaN	(18,99-23,01)	(18,76-23,07)	(18,76-23,07)	(16,82-22,38)	(20,81-27,18)	(20,81-27,18)	(19,59-25,37)	(15,21-21,03)	(15,21-21,03)	Total
2	permanece igual	0.546111	0.536461	0.536461	0.551849	0.533913	0.533913	0.540612	0.538903	0.538903	Total
3	NaN	(52,17-57,03)	(51,1-56,18)	(51,1-56,18)	(51,7-58,62)	(49,73-57,02)	(49,73-57,02)	(50,63-57,45)	(50,33-57,41)	(50,33-57,41)	Total
4	empeora la salud	0.244624	0.255253	0.255253	0.253662	0.227603	0.227603	0.235962	0.281747	0.281747	Total
5	NaN	(22,39-26,66)	(23,37-27,81)	(23,37-27,81)	(22,42-28,56)	(19,84-25,97)	(19,84-25,97)	(20,76-26,69)	(25,08-31,49)	(25,08-31,49)	Total
6	mejora la salud	0.285829	0.207947	0.207947	0.225585	0.298627	0.298627	0.348936	0.112922	0.112922	16-29
7	NaN	(23,63-34,11)	(16,24-26,22)	(16,24-26,22)	(16,8-29,59)	(22,25-38,79)	(22,25-38,79)	(27,16-43,51)	(7,51-16,64)	(7,51-16,64)	16-29

Solución del problema 2

... a esto:

In [22]: `df[0][0]`

Out [22]:

	edad	autop	Medida	est	lim_inf	lim_sup
0	Total	mejora la salud	total_M2-M1_Porc	20.9265	18.99	23.01
1	Total	permanece igual	total_M2-M1_Porc	54.6111	52.17	57.03
2	Total	empeora la salud	total_M2-M1_Porc	24.4624	22.39	26.66
3	16-29	mejora la salud	total_M2-M1_Porc	28.5829	23.63	34.11
4	16-29	permanece igual	total_M2-M1_Porc	47.0377	41.43	52.72
...	...	...	...	...	...	...
130	45-65	permanece igual	mujeres_M4-M3_Porc	59.8409	54.14	65.28
131	45-65	empeora la salud	mujeres_M4-M3_Porc	25.2048	20.58	30.47
132	65+	mejora la salud	mujeres_M4-M3_Porc	23.2545	16.34	31.97
133	65+	permanece igual	mujeres_M4-M3_Porc	51.9349	43.51	60.25
134	65+	empeora la salud	mujeres_M4-M3_Porc	24.8106	18.45	32.49

135 rows × 6 columns

Resultados: Gráficas

Para cada variable se almacenan en gráficas[*l*][*i*] siguiendo la notación anterior. Estas gráficas se almacenarán en la Web de la ESSOC.

Atributos importantes para la creación de gráficas:

- alt.Chart
- Facet
- selection
- alt.hconcat

Resultados: Gráficas

Ejemplo:

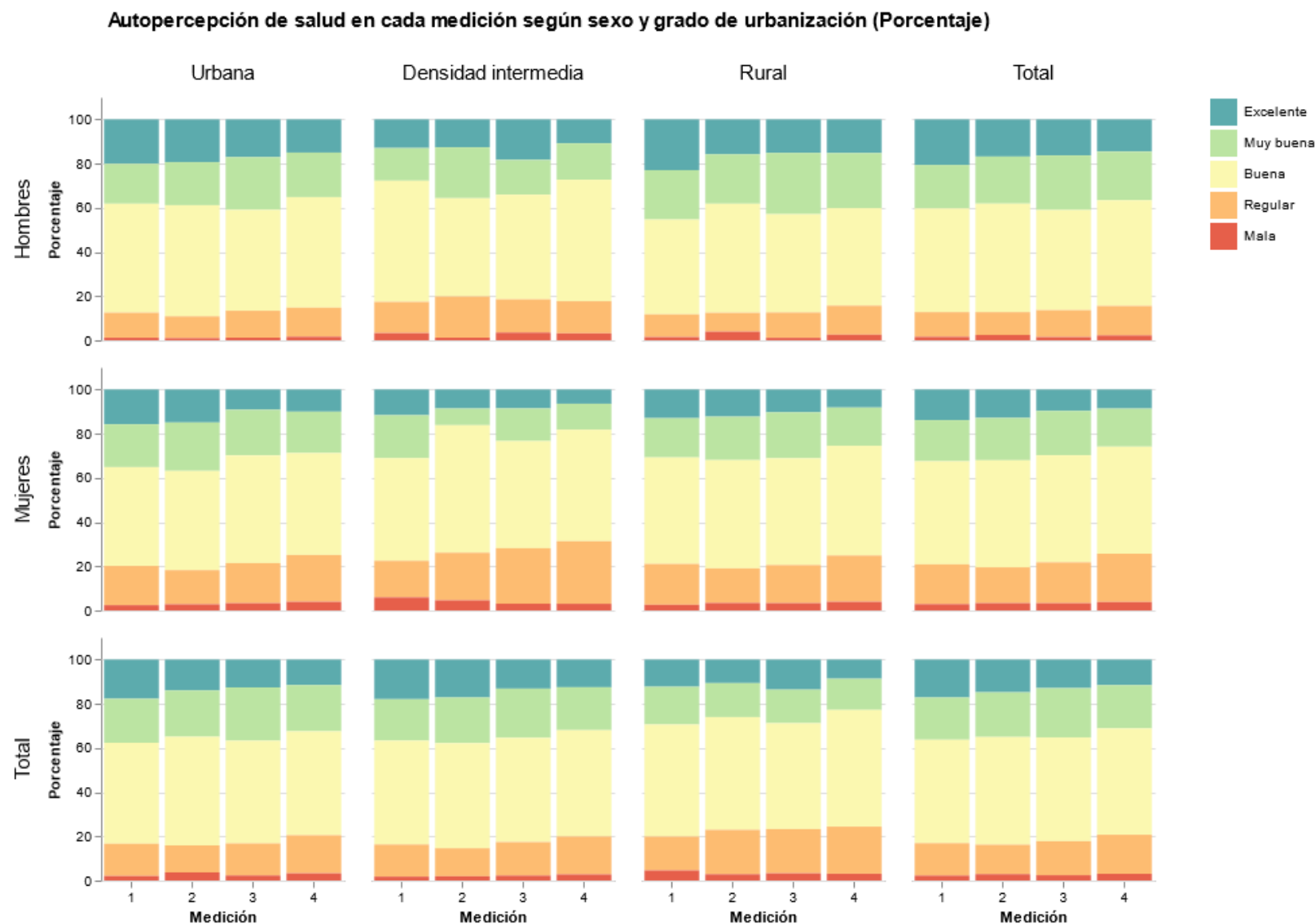
```
selection = alt.selection_multi(fields=['autop'], bind='legend')

bar_chart = alt.Chart(df_long_dif).mark_bar(
).encode(
  x=alt.X('Medición:N', title='Medición', axis=alt.Axis(labelAngle=0)),
  y=alt.Y('est:Q', title='Porcentaje'),
  order='order', # this controls stack order
  color=alt.Color('autop:N', scale=alt.Scale(scheme='spectral', reverse=True),
                  # optional: make color order in legend match stack order
                  sort=alt.EncodingSortField('order', order='descending'),
                  legend=alt.Legend(title="",
                                   symbolSize=400, symbolType='square', labelLimit=0)),
  tooltip=[alt.Tooltip('est', title='% estimado'), alt.Tooltip('lim_inf', title='Límite inferior'),
           alt.Tooltip('lim_sup', title='Límite superior')]
  # opacity=alt.condition(selection, alt.value(1), alt.value(0.2))
).add_selection(selection).transform_filter(selection).properties(width=180, height=180)
).facet(title='Título', row=alt.Row('Sexo:N', title=None, header=alt.Header(labelFontSize=15)),
        column=alt.Column('edad:N', title=None, header=alt.Header(labelFontSize=15)))
```

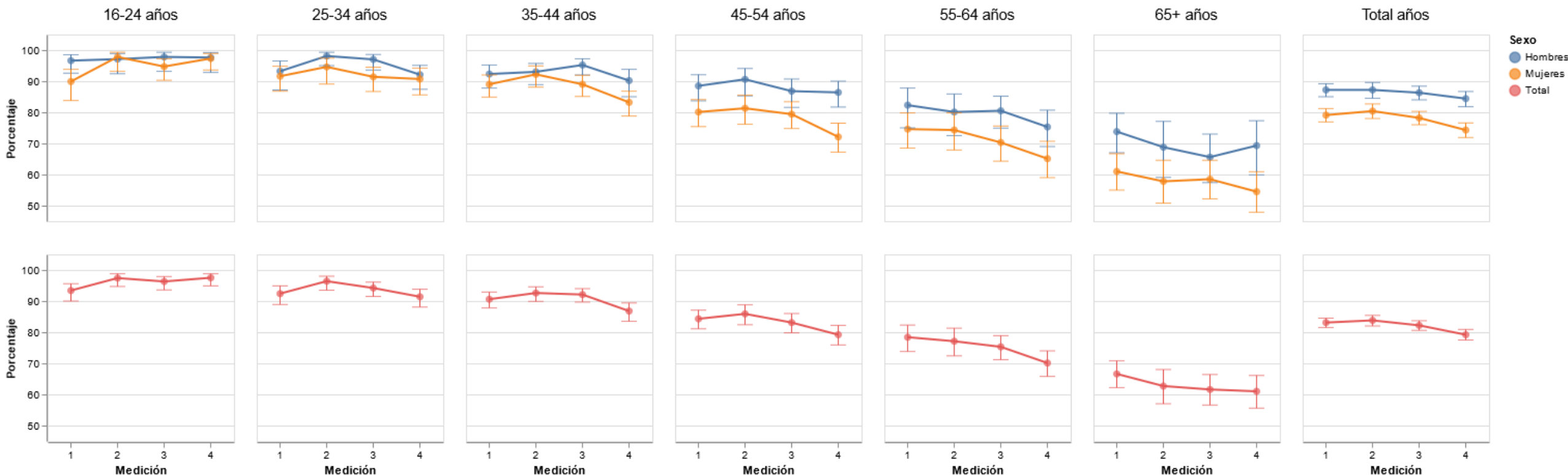
Resultados: Gráficas

```
leyenda = alt.Chart(df_long_dif).mark_text().encode(
    color=alt.Color('autop:N', scale=alt.Scale(scheme='spectral', reverse=True),
    # optional: make color order in legend match stack order
    sort=alt.EncodingSortField('order', order='descending'),
    legend=alt.Legend(title="",
    symbolSize=400, symbolType='square', labelLimit=0, offset=-170)),
    opacity=alt.value(0.2)
).add_selection(selection
)

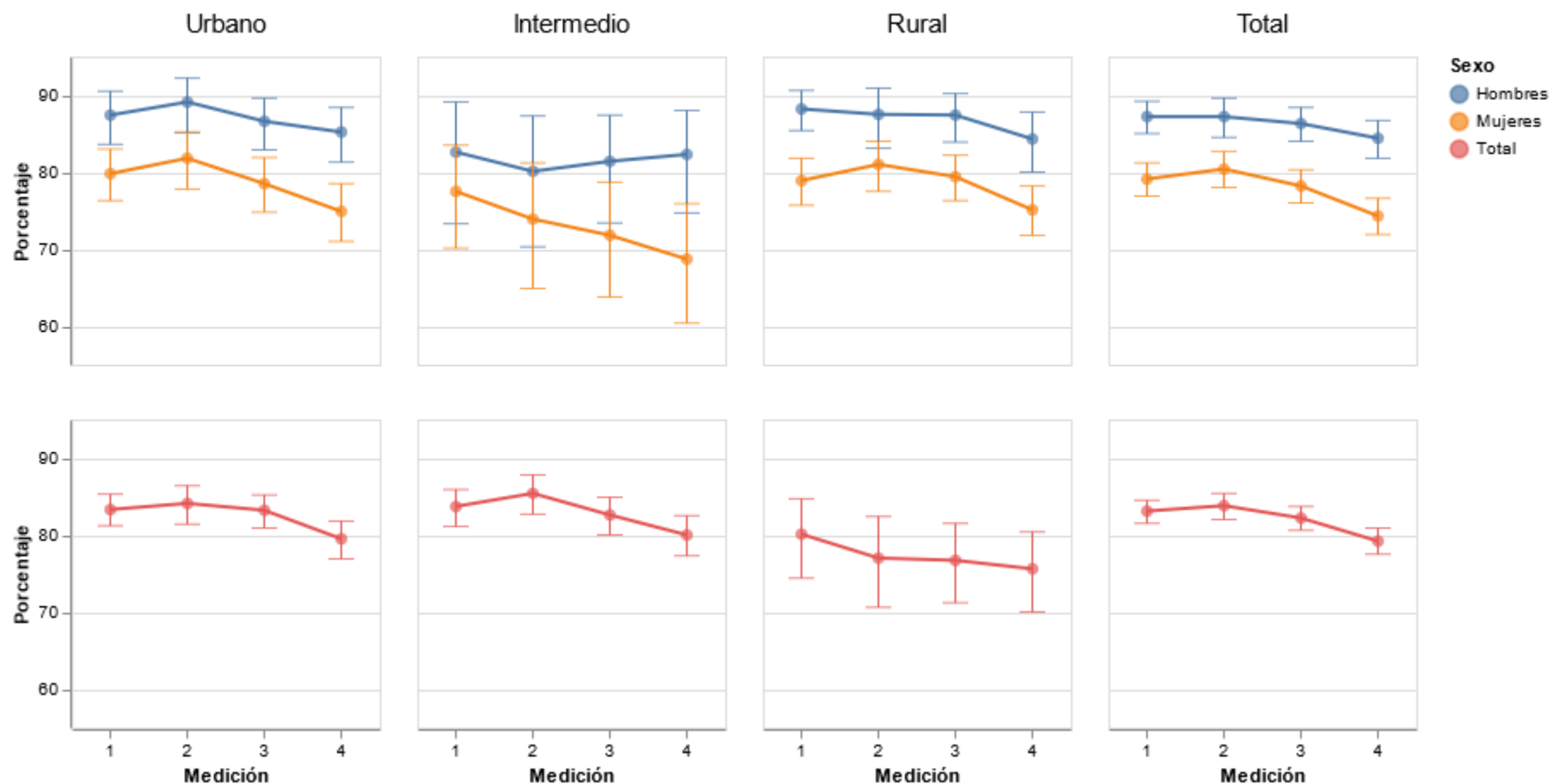
long_dif_P = alt.hconcat(bar_chart, leyenda).configure_view(strokeWidth=0).configure_title(
    color='black',
    dy=-20,
    dx=70,
    fontSize=15)
```



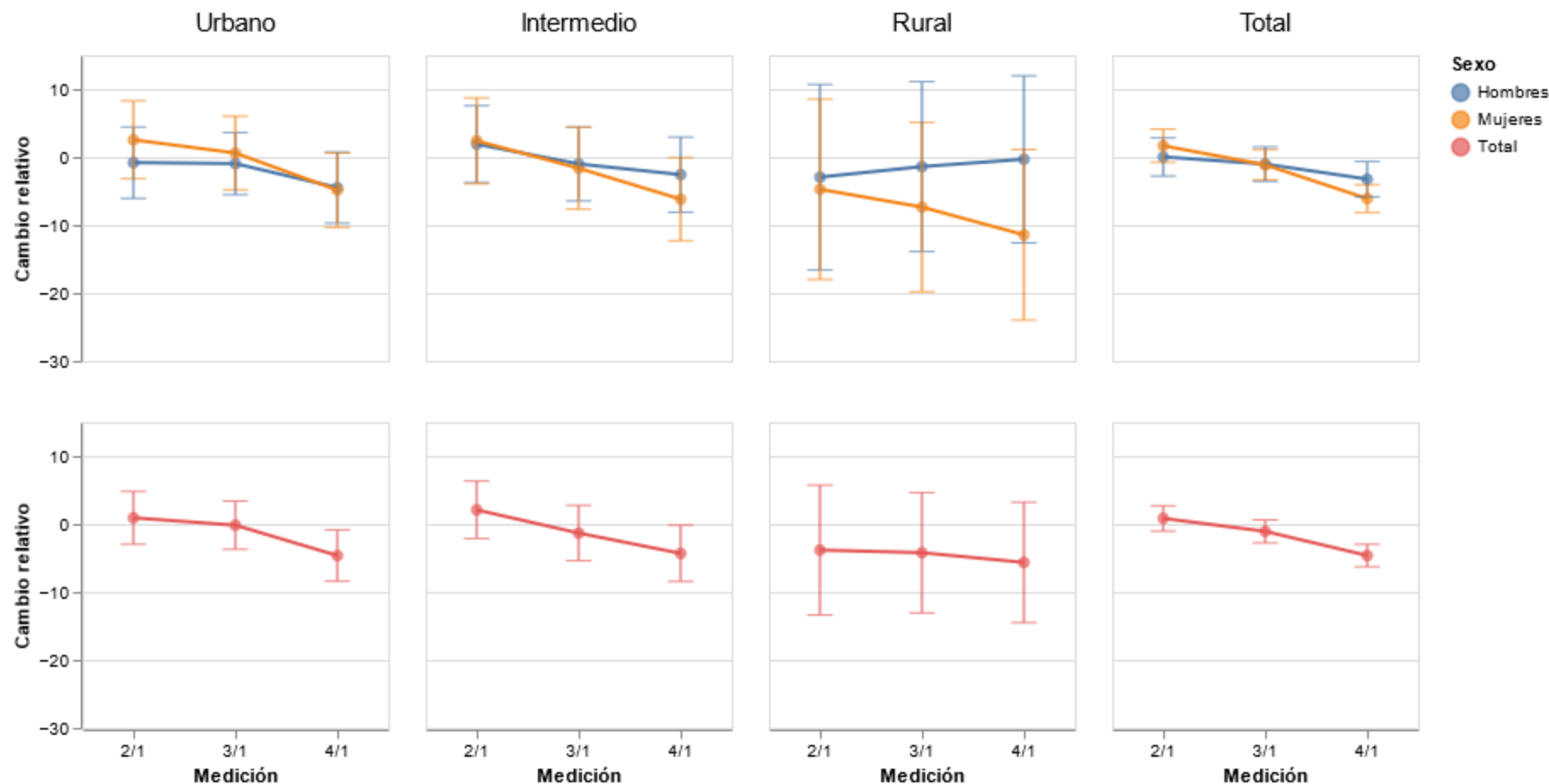
Autopercepción de salud excelente o muy buena en cada medición según sexo y edad (porcentaje e intervalos de confianza al 95%)



Autopercepción de salud excelente o muy buena en cada medición según sexo y grado de urbanización (porcentaje e intervalos de confianza al 95%)



Cambio en autopercepción de salud excelente o muy buena en cada medición según sexo y grado de urbanización, intervalos de confianza al 95% (Cambio relativo)



Automatización de la creación de gráficas

Objetivo: crear todas las gráficas para las más de 1000 variables de manera automática.

Método: Usar diccionarios en Python para crear código genérico e iterar para todas las variables.

Ejemplo: maintitle = { '16': 'Autopercepción de salud',
 '17': 'Autopercepción de salud mental',
 '18': 'Sentimiento de depresión' }.

Resultados: Web de la ESSOC

Encuesta Sanitaria y Social: visualización resultados descriptivos

Descargar todas las gráficas y tablas en .PDF

Seleccione la/s variable/s resultado

Características del hogar y la vivienda

Uso del tiempo y convivencia

Salud y Bienestar emocional

Hábitos y estilo de vida

Situación económica y características sociodemográficas

Seleccione el/los resultado/s descriptivo/s

Todas

Transversal variables originales

Transversal variables dicotomizadas

Transversal cambio autopercepción

Transversal cambio brecha género

Longitudinal diferencia

Longitudinal diferencia brecha género

Seleccione la/s variable/s de estratificación

Todas

Grupo de edad

Provincia

Grado de urbanización

Visualizar

Resultados: Web de la ESSOC

Encuesta Sanitaria y Social: visualización resultados descriptivos

Descargar todas las gráficas y tablas en .PDF

Seleccione la/s variable/s resultado

Características del hogar y la vivienda

Todas

Variable Ejemplo: Opción 1 Opción 2

Opción 3 Opción 4 Opción 5

Superficie:

Instalaciones:

Seleccione el/los resultado/s descriptivo/s

Todas

Transversal variables originales

Transversal variables dicotomizadas

Transversal cambio autopercepción

Transversal cambio brecha género

Longitudinal diferencia

Longitudinal diferencia brecha género

Seleccione la/s variable/s de estratificación

Todas

Grupo de edad

Provincia

Grado de urbanización

Visualizar

Encuesta Sanitaria y Social: visualización resultados descriptivos

Descargar todas las gráficas y tablas en .PDF

Variable descriptiva: Autopercepción de salud

Resultado descriptivo: Transversal Variables Originales

Grupo de edad

Estimador: Porcentual Total

Visualización: Gráfica Tabla

Descargar

Autopercepción de salud en cada medición según sexo y edad (N)

Encuesta Sanitaria y Social de Hogares: análisis y visualización mediante R y Python

Jorge Hidalgo Calderón

23

Conclusión

Altair permite crear visualizaciones personalizadas e interactivas.

Estas gráficas permiten interpretar la cantidad ingente de información de las más de 1000 variables y más de 6000 tablas producidas por la ESSOC.

La Web de la ESSOC con todos los resultados presentados en esta presentación se encontrará en:

www.researchproject.com/ESSOCgeneral .

Referencias

- Sánchez-Cantalejo, C., Rueda, M. M., Saez, M., Enrique, I., Ferri-García, R., De la Fuente, M., Castro-Martín, L., ... & Cabrera-León, A. (2021). Impact of COVID-19 on the health of the general and more vulnerable population and its determinants: Health care and social survey-ESSOC, study protocol.
- Luis Castro-Martín, María del Mar Rueda, Andrés Cabrera, Carmen Sánchez-Cantalejo, Ramón Ferri-García, Jorge Hidalgo. Reweighting with machine learning techniques in panel surveys. Application to the Health Care and Social Survey. The 19th Conference of the Applied Stochastic Models and Data Analysis International Society ASMDA2021
- <https://altair-viz.github.io/>



Islas Canarias
Del 15 al 19 de noviembre de 2021

iGracias!



Escuela Andaluza
de Salud Pública
Consejería de Salud y Familias



UNIVERSIDAD
DE GRANADA

