



PRAC1:

Visualización de datos.

Creación de la
visualización: Factores
determinantes en
infarto de miocardio.

Pablo Santos Ramos

09 de junio de 2021

1. PRESENTACIÓN DE LA VISUALIZACIÓN

1.1. *Título de la visualización.*

Visualización de los distintos factores determinantes en un infarto de miocardio.

1.2. *URL del repositorio.*

La tarea completa junto con el informe se puede encontrar en el repositorio web de GitHub:

https://github.com/PabloSantos11/PRAC_FINAL_Visualizacion_de_datos

1.3. *Descripción breve de la visualización.*

Existen multitud de factores que pueden afectar y hacer más propensos a las personas a sufrir un infarto de miocardio. En esta visualización estudiaremos a diversos pacientes que han sufrido un infarto o que simplemente tienen una angina de pecho que pueda derivar en uno.

La siguiente visualización comenzará con algunos gráficos de las variables a priori más relevantes, con su posterior comparación entre ellas, de manera que podamos extraer ideas rápidas y simples con solo verlos. Seguidamente se analizarán algunas variables en función del número de infartos sufridos por esa persona para, finalmente, acabar en una visualización interactiva con la que podamos explorar y encontrar distintos patrones según vayamos interactuando con ella.

2. EXPLICACIÓN DE LA VISUALIZACIÓN

Se ha escogido esta visualización debido a que el infarto de miocardio es una enfermedad cardíaca que afecta a millones de personas en el mundo y a miles diariamente. Encontrar algunos patrones o tendencias en personas antes de que estos se produzcan puede salvar muchas vidas. En ocasiones, tan solo con el transcurso de unas horas la enfermedad puede ser extremadamente dañina o simplemente un dolor puntual.

Al visualizar diversas variables determinantes, podemos encontrar rápidamente conclusiones (que aunque superficiales y no exactamente corroboradas por un análisis estadístico) que nos ayuden a diagnosticar mucho antes si una persona puede o va a sufrir un infarto de miocardio.

Esta visualización principalmente va dedicada a trabajadores del sector Médico y a gente interesada en la Cardiología. Con esta visualización vamos a poder responder a multitudes de preguntas, como por ejemplo:

- ¿Qué relevancia tienen las anginas de pecho a la hora de sufrir un infarto?
- ¿Qué tipo de anginas de pecho es el más común entre la población?
- ¿Influye la edad o el sexo en la cantidad de infartos sufridos?
- ¿Se puede mejorar la hipertensión según avancen los años de diagnóstico?
- ¿Cuál es el daño más común que puede causar un infarto?

- Al sufrir un infarto, ¿Cuánto tiempo tardan los pacientes en acudir al hospital?

Como vemos, existen muchas preguntas a las que responder, incluso se podrían añadir algunas más, no obstante, con la siguiente visualización vamos a poder resolverlas con un simple vistazo rápido, por lo que se pueden explicar otras cuestiones sin necesidad de que aparezcan en las preguntas.

3. DESCRIPCIÓN TÉCNICA

Los datos del data set escogido eran todos numéricos, de manera que para saber el significado de cada número debemos mirarlo en el documento de GitHub ‘Explicación de variables’. Debido a esto, se ha realizado una transformación a diversas variables para poder entender correctamente las visualizaciones, ya que el objetivo principal de una visualización es mostrar la información completa y simple, de manera que podamos interpretar claves de los datos sin necesidad de analizarlos.

Las variables transformadas han sido el número de infartos, para poder graficar y rellenar gráficos de barras; la hipertensión, de manera que podamos entender en que fase nos encontramos; el daño causado por el infarto, ya que cada número significaba un *outcome* y el tiempo de llegada al hospital.

Todas estas transformaciones y algunas de las visualizaciones se han realizado con el software R studio, ya que es el utilizado a lo largo del curso y ofrece herramientas simples y rápidas para las visualizaciones. Se han necesitado 3 librerías externas para la ejecución del código:

- Librería ggplot2
<https://ggplot2.tidyverse.org/>
- Librería scales
<https://cran.r-project.org/web/packages/scales/index.html>
- Librería openxlsx
<https://cran.r-project.org/web/packages/openxlsx/index.html>

Para finalizar, se ha creado una gran visualización interactiva con la herramienta Flourish, de manera que se pueda cambiar los factores y las variables para intuir y observar nuevos patrones.

El lenguaje utilizado es R bajo la licencia: Licensed under CC-BY 4.0 2018–2020 by The Carpentries. El software Flourish al ser ‘Public’ no requiere licencia adicional.

El data set se ha obtenido de la página UCI ‘Machine Learning Repository’ y se puede acceder mediante el link:

<https://archive.ics.uci.edu/ml/datasets/Myocardial+infarction+complications>

Los creadores de la base de datos junto con sus estudios y universidades son:

Creators of database:

S.E. Golovenkin, V.A. Shulman, D.A. Rossiev, P.A. Shesternya, S.Yu. Nikulina, Yu.V. Orlova:
Professor V.F. Voyno-Yasenetsky Krasnoyarsk State Medical University;

A.N. Gorban, E.M. Mirkes: University of Leicester.

Database donator:

E.M. Mirkes, School of Mathematics and Actuarial science, University of Leicester, Leicester, LE1
7RH, UK, em322 '@' le.ac.uk

4. VISUALIZACIÓN DE DATOS

En primer lugar, debemos seleccionar las variables que utilizaremos en la visualización, para ello cargamos el conjunto de datos total 'Myocardial' y con R extraemos otro data frame con los factores determinantes.

A continuación, se visualizan dos gráficos sobre las anginas de pecho, que son deterioros u obstrucciones de las arterias del corazón que posteriormente pueden derivar en infartos. En el siguiente gráfico observaremos, por un lado, el tiempo en años desde que se ha detectado la angina al paciente y por el otro el tipo de angina del que se trata. Según tenga una Clase mayor, la angina es más propensa a aparecer sin necesidad de que la persona se mueva o haga ejercicio. Los gráficos explicativos de la angina son:

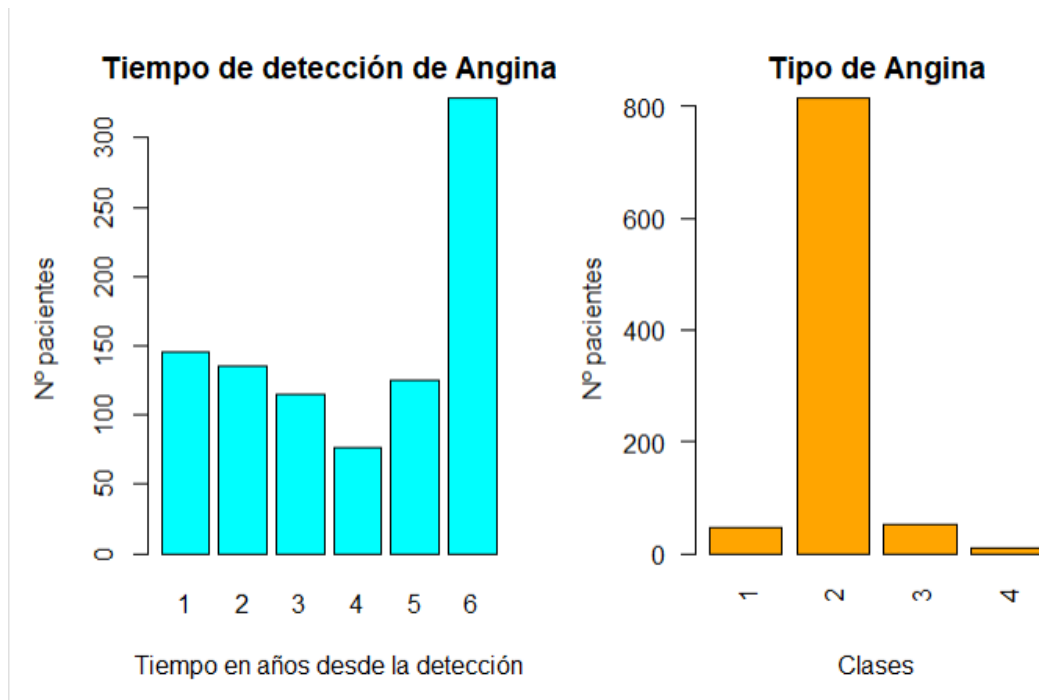


Figura 1. Gráficos explicativos de angina de pecho.

Como podemos observar la mayor parte de las anginas de pecho llevan diagnosticadas 6 años o más y la clase de angina más común es la angina de Clase 2, caracterizada por el dolor de pecho simplemente con andar rápido o subir escaleras o cuestas.

Nos centramos ahora en el número de infartos en función de otros atributos como son la edad o el sexo. En nuestro conjunto de datos existe la variable 'N_infartos' que nos indica cuántos infartos ha sufrido el paciente excepto si el campo es '0' que indica simplemente la presencia o detección de una angina de pecho.

En función del sexo del paciente podemos realizar la siguiente visualización:



Figura 2. Número de infartos en función del sexo del paciente.

No podemos sacar grandes conclusiones ya que no es una visualización muy significativa, sin embargo, nos puede ayudar en futuras visualizaciones a encontrar tendencias relacionadas. En este gráfico observamos ligeramente una predominancia de los hombres a la hora de tener más infartos.

Relacionamos ahora el número de infartos con la edad:

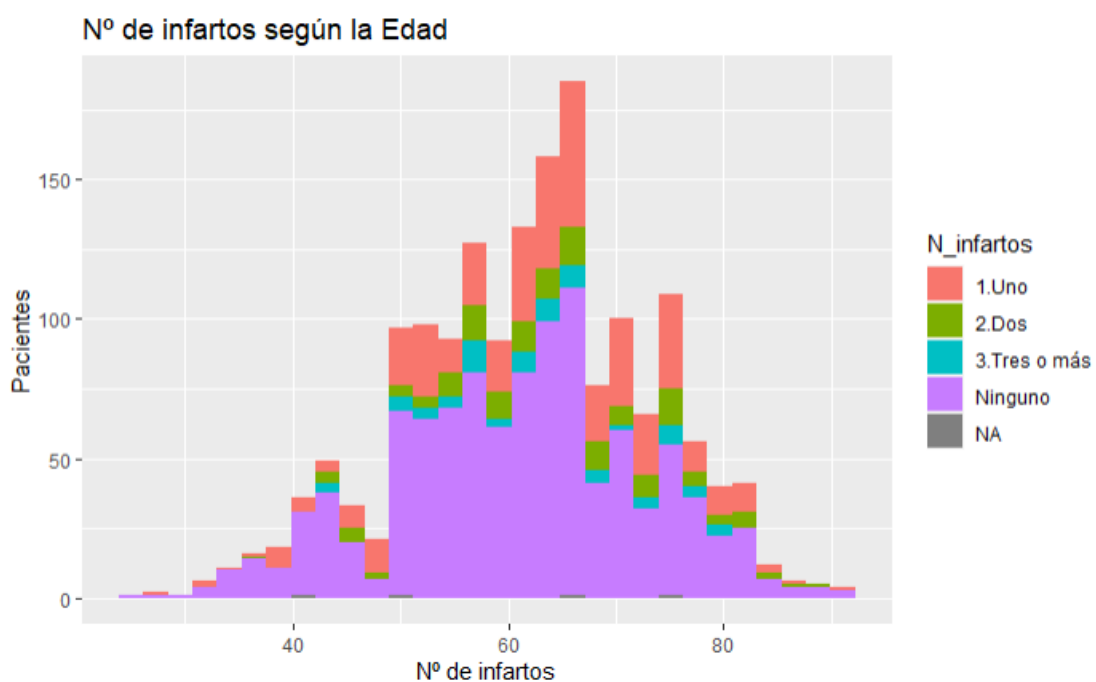


Figura 3. Número de infartos en función de la edad del paciente.

Observamos en este caso que la predominancia en todas las edades es la de ningún infarto, es decir, simplemente se les ha diagnosticado una angina de pecho. No obstante, se observa la predominancia de infartos en torno a los 65 años, siendo en ocasiones antes de los 60 cuando se producen hasta 3 infartos por paciente.

Es clara la poca cantidad de infartos en personas jóvenes y el gran incremento que se produce una vez los pacientes superan los 50 años.

Terminando las visualizaciones con Rstudio, realizamos un gráfico que compara la fase de hipertensión en la que se encuentra el paciente con la duración de esta. Para ello ha sido necesario transformar las variables numéricas del data set a categóricas de manera que se pueda obtener un gráfico como los anteriores.

La visualización de la variable hipertensión es la siguiente:

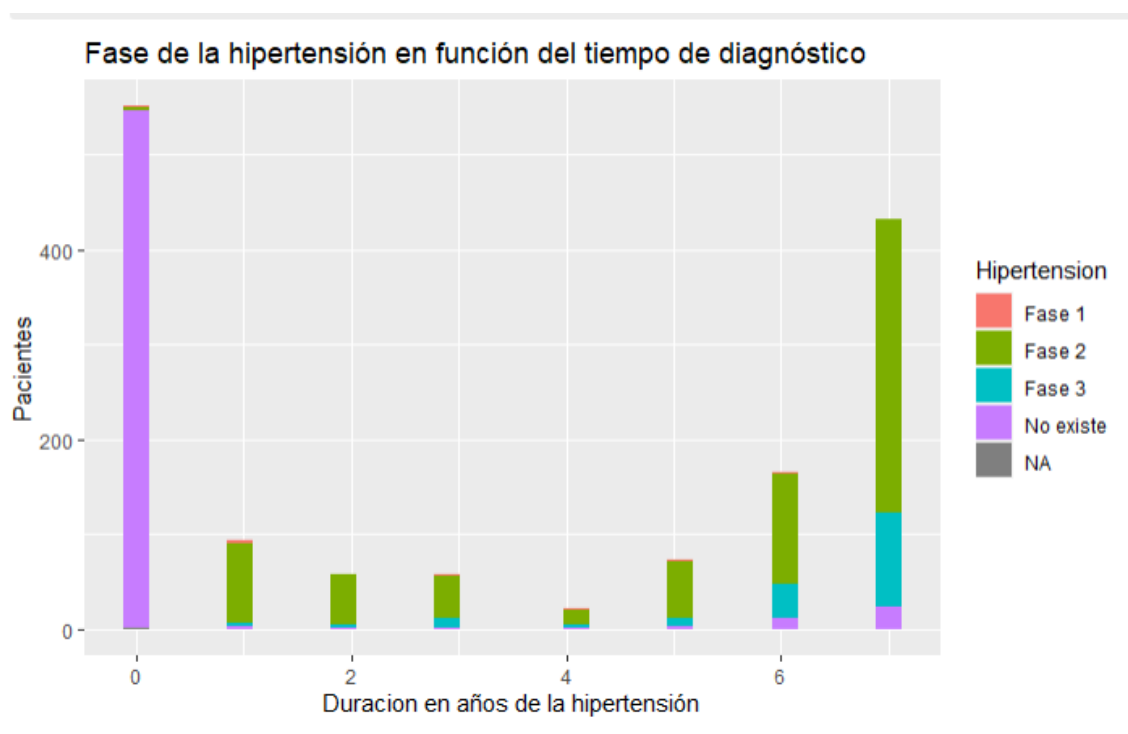


Figura 4. Fase de la hipertensión según el tiempo diagnosticada.

De este gráfico podemos extraer que, según los pacientes lleven más años con hipertensión diagnosticada, la fase de esta es mayor. Por ejemplo, cuando solo lleva dos años diagnosticada vemos que casi todo es fase 2, mientras que cuando los pacientes llevan diagnosticados 6 ó 7 años la fase 3 empieza a notarse. Incluso se puede ver que si llevan muchos años con hipertensión desaparece en ocasiones (parte morada).

Procederemos ahora a la creación de un data frame con distintas variables relevantes con las que haremos una visualización interactiva en Flourish. Una vez creado el data frame lo

exportaremos en formato .xlsx para poder importarlo directamente a la visualización. El link del repositorio donde se encuentra esta visualización es el siguiente:

<https://app.flourish.studio/projects?folder=%1EPRA%20FINAL%20visualizacion&filter=all>

No obstante, todos estos links, junto con el código R los encontraremos en GitHub.

Una vez nos encontramos dentro de Flourish podemos escoger diferentes pestañas con las que podemos agrupar según diferentes atributos, sombrear e incluso cambiar el tamaño. Para que se puedan ver la cantidad de combinaciones disponibles, se presentarán algunas de ejemplo en este informe:

En el primer ejemplo vamos a agrupar por el tiempo de llegada al hospital, es decir, desde que se produce el infarto o la angina de pecho, cuánto tarda el paciente en acudir a urgencias. Después sombrearemos en función del número de infartos sufridos y compararemos ambos sexos. El resultado final es:

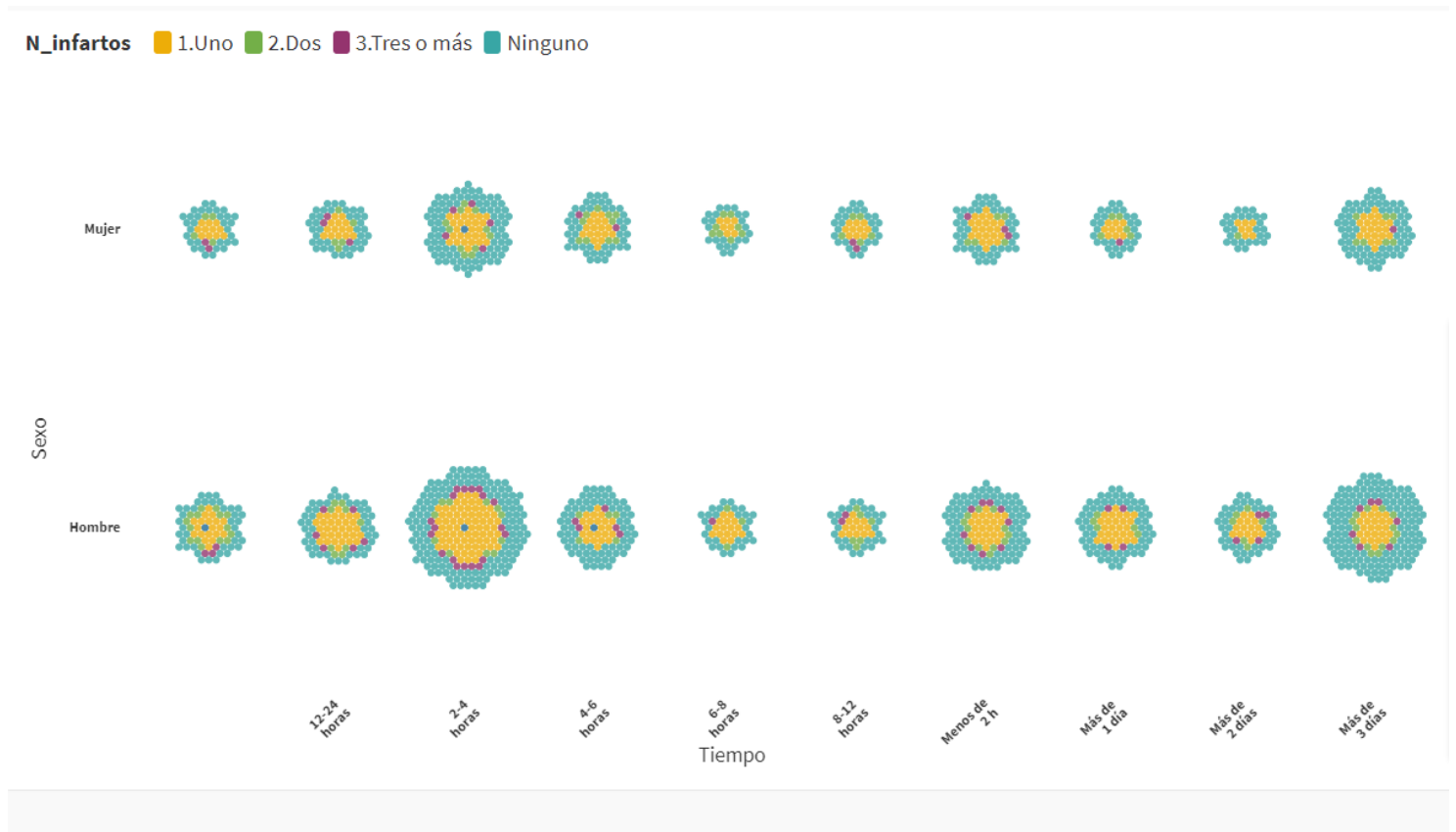


Figura 5. Tiempo de llegada en función de infartos y sexo.

Observamos que existen más infartos que llegan al hospital entre las 2-4 horas después de producirse y a los 3 días de producirse. Como anteriormente hemos visto con menos detalle, existe una ligera predominancia en infartos de los hombres.

En el siguiente ejemplo podemos cambiar el formato de la visualización y comparar el número de infartos en función de la angina de pecho detectada, de manera que podamos ver cuál de estas anginas es el indicador de posible infarto:

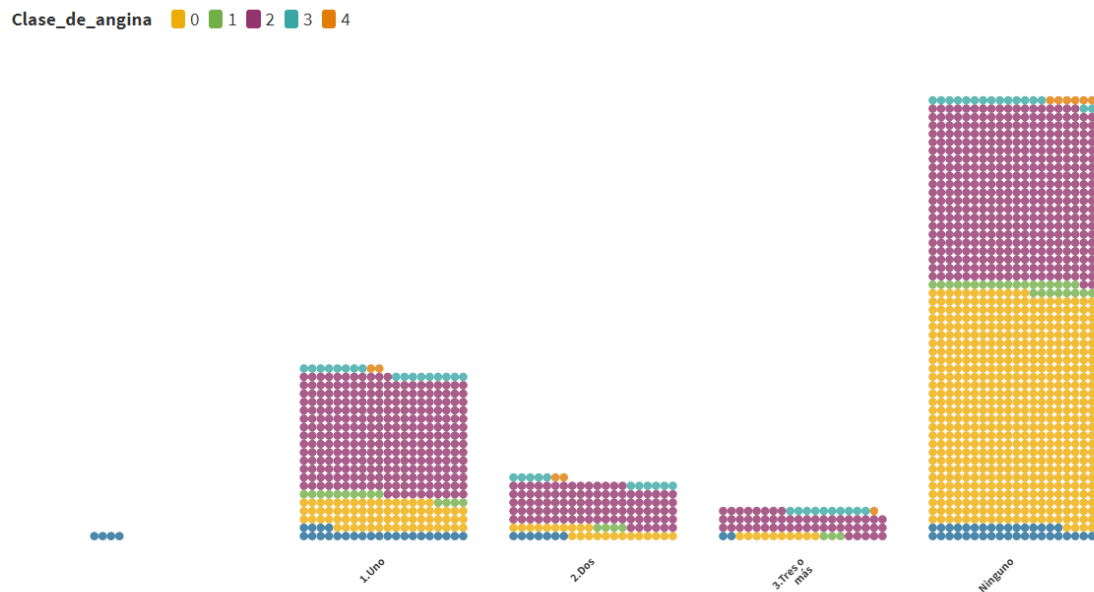


Figura 6. Clase de angina en función del número de infartos.

De esta visualización claramente podemos observar que los pacientes cuya angina de pecho es de clase 2 son más propensos a sufrir un infarto (predominancia del morado). Según el número de infartos va aumentando la predominancia de la clase 2 de angina va aumentando con este. Los casos de pacientes de Clase 3 y 4 son similares en todos los casos, incluso en los que no han tenido infarto y solo se les ha diagnosticado la angina.

Un último ejemplo puede ser la combinación de estas dos visualizaciones anteriores de manera que podamos obtener una especie de tabla de la que podamos analizar, pulsando en cada círculo paciente por paciente. En el informe no se puede observar muy bien, no obstante, se adjunta una captura para que se pueda replicar:

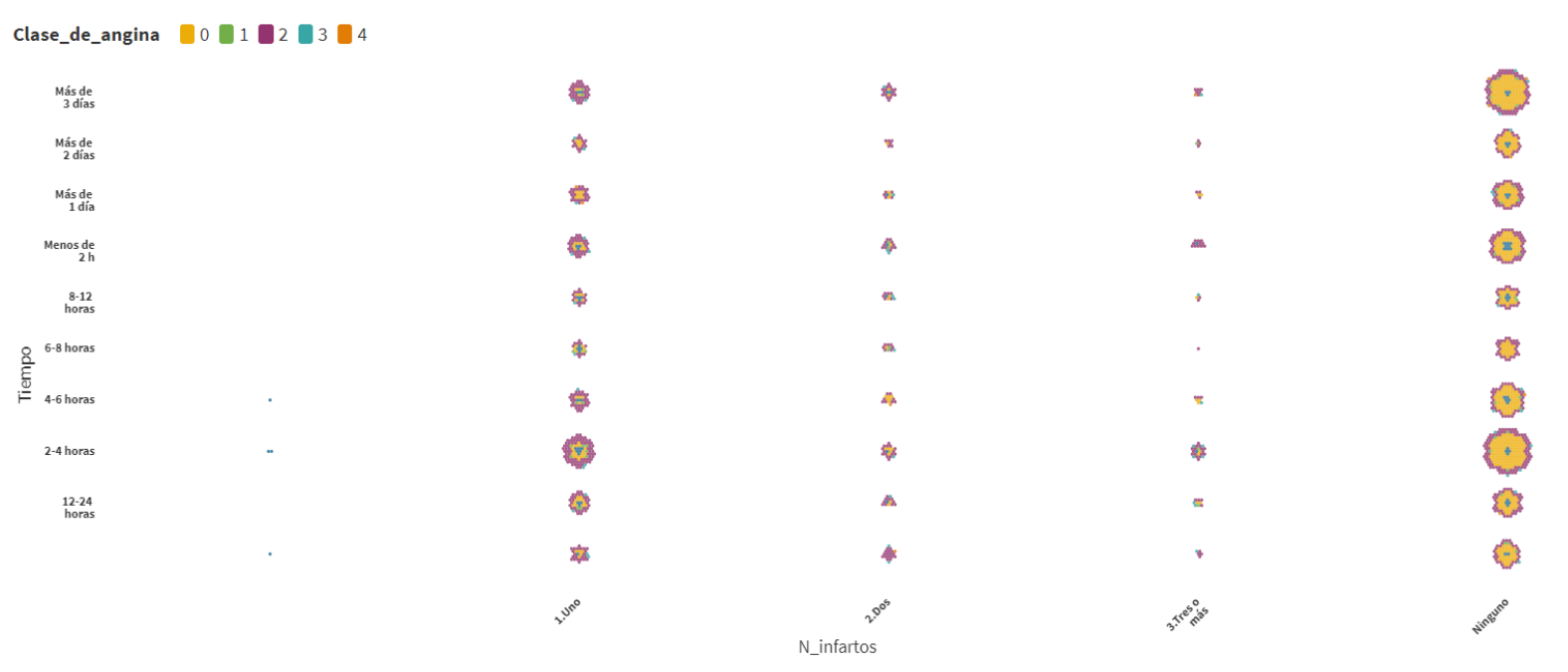


Figura 7. Tiempo de llegada en función de clase de angina y del número de infartos.

Con el link propuesto anteriormente, se puede acceder libremente a la herramienta Flourish y se puede ‘jugar’ con los distintos parámetros escogidos de la manera que uno desee.

5. ENTREGA DE LA TAREA

La tarea completa junto con el informe se puede encontrar en el repositorio web de GitHub:

https://github.com/PabloSantos11/PRAC_FINAL_Visualizacion_de_datos

En él encontraremos los data sets mencionados en el informe, el código de R junto con su respectivo HTML para evitar ejecuciones innecesarias, el link hacia la visualización interactiva y todos los documentos necesarios para la correcta visualización de los factores determinantes en el infarto de miocardio.