

Situación problema: Kit diagnóstico para la detección
temprana de cáncer de colon en adultos jóvenes

Luis Humberto Sánchez Vaca
A01638029

Óscar Miranda Escalante
A01630791

José de Jesús Gutiérrez Aldrete
A01637812

Ana Ivette Farías Rodríguez
A01637864

José Gerardo Villanueva Barceló
A01636721

1 de mayo de 2020

Introducción

Para poder comenzar con la investigación realizada, primeramente se dejaron claro conceptos básicos que nos ayudan a comprender de manera superficial el tema a tratar para ir adentrándonos poco a poco, comenzando por conocer la definición formal de colon y las funciones de este órgano; “el colon es la parte final de nuestro sistema digestivo, se le conoce comúnmente como intestino grueso” [1], su propósito es vital para el funcionamiento adecuado de la vida de un ser humano y dichas funciones son:

- En el colon se produce la absorción de la mayor parte de agua contenida, de manera que el material que sale al final es fundamentalmente sólido, poco hidratado, y es gracias a esto que el organismo puede mantener un estado de hidratación adecuado.
- Absorción de vitaminas: A lo largo del colon se absorben diversos componentes del complejo de vitaminas importantes para el desarrollo neurológico, de la visión y la adecuada formación de los glóbulos rojos.
- En el colon tiene lugar la recirculación de componentes biológicos, tales como algunos ácidos grasos de cadena corta y la urea.

Igualmente, tenemos que entender qué es el cáncer, pues esta enfermedad es en lo que se centra nuestra investigación; La Sociedad Española de Oncología Médica define el cáncer como “un grupo numeroso de enfermedades que se caracterizan por el desarrollo de células anormales, que se dividen, crecen y se diseminan sin control en cualquier parte del cuerpo” [2], el cáncer puede empezar casi en cualquier lugar del cuerpo humano, el cual recordemos que está formado de montones de células, las cuales por lo general crecen y se dividen para formar nuevas, pero en el cáncer, este proceso se descontrola lo que hace que las células viejas o dañadas sobreviven cuando deberían morir, y células nuevas se forman cuando no son necesarias. Estas células adicionales pueden dividirse sin interrupción y pueden formar masas que se llaman tumores [2, 3].

Podemos encontrar algunas conductas de riesgo, esto quiere decir, acciones que conlleven a padecerlo [4], como por ejemplo:

- Falta de actividad física habitual
- Alimentación pobre en frutas y verduras
- Alimentación pobre en fibra y rica en grasas, o rica en carnes procesadas
- Sobrepeso y obesidad
- Consumo de alcohol
- Tabaquismo

Aún así, no todo es por estilo de vida, pues esta enfermedad la padecen en un 90 % los adultos mayores a 50 años y también puede ser otras enfermedades como enfermedades inflamatorias intestinales o genética.

Antecedentes

Con el transcurso del tiempo el cáncer colorrectal se posicionó en la sexta enfermedad tumoral, en conjunto a ser la de mayor incidencia en el sistema digestivo. Hablamos de una patología con un crecimiento exponencial en México, convirtiéndose así en la segunda causa principal de muerte de cáncer y el tercero más frecuente en patología oncológica. Afecta por igual a mujeres y hombres con una incidencia en la edad entre 60 y 65 años, pero en los últimos años se han presentado un aumento de casos en personas entre los 20 y 40 años.[6]

El IMSS comunicó que “el 36 % de los casos nuevos de cáncer [son] de colon y otro 37 % de los casos nuevos de cáncer de recto; por su parte, la Secretaría de Salud reportó el 15 % de los casos de cáncer [son] de colon y el 18 % de (...) de recto, equivalentes a una tercera parte del total de

casos registrados en México”. Mediante los resultados de casos nuevos por cáncer de colon y recto por grupo de edad, México, 2002, nos indica que surgieron 1254 casos en hombres y 1269 casos en mujeres en el cancer de colon, respecto al cáncer en el recto 674 corresponden en hombres y 594 en mujeres. En total se presentaron 3791 nuevos casos en el respectivo año. Encontramos una relación entre la edad y el padecimiento al compararlo con el panorama internacional, ya que en el 2002 se registraron 550,465 casos nuevos en hombres y 472,687 en mujeres. Estos datos indican una incidencia de padecimiento a partir de los 45 años. Por lo tanto podemos deducir que los casos nuevos son directamente proporcional a la edad. Respecto a la tasa de mortalidad nos indica que el cáncer de colon tiene una mayor frecuencia que el cáncer de recto.[6]

Analizando los datos de pacientes de nuevo ingreso del Instituto Nacional de Cancerología (IN-CAN) durante el 2016 se obtuvieron los siguientes resultados. En el apéndice del presente reporte podrá encontrar la **tabla con los casos de cáncer colorrectal de la población mexicana**[5]

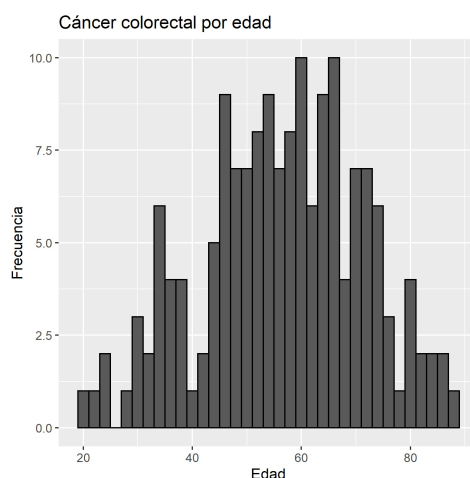


Figura 1: Casos de cáncer de colon por edad

En esta grafica podemos observar la frecuencia de casos de cáncer colorrectal por edades, es notorio que la frecuencia de casos aumenta a partir de los 40 años y disminuye antes de los 80, mediante estos datos del INCAN podemos compararlo con índices anteriores los cuales mencionan una incidencia de padecimiento a partir de los 45 años y que este mismo está correlacionado con el número de casos y edad. Ante la preocupación por la incidencia de casos en adolescentes y adultos menores a 45 años se encontró que se han presentado 29 casos, lo cual también comprueba que han surgido casos a estas tempranas edades.

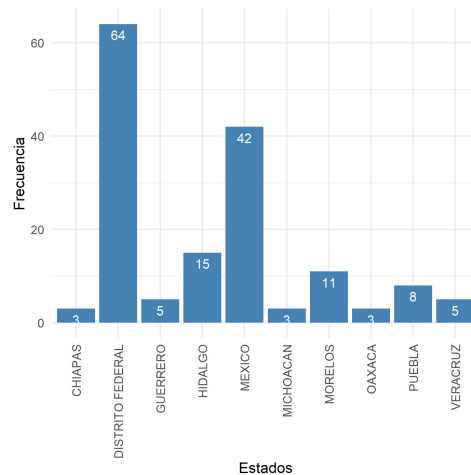


Figura 2: Cáncer colorrectal por estado[5]

Como podemos observar en esta gráfica, los principales 3 estados con mayor frecuencia en México son Ciudad de México(Distrito Federal), México e Hidalgo. Según el reporte de Epidemiología del cancer y colon “Del total de defunciones registradas para este año (58,612), el 4.48 % (2,602) correspondieron a cáncer de colon (2,178) y cáncer de recto (424), con tasas de 2.1 y 0.4 por 100,000 habitantes, respectivamente. Las entidades federativas con mayores tasas de mortalidad fueron Distrito Federal y Chihuahua” [6]. Estos datos nos indican que la incidencia se puede deber a los factores ambientales, la correlación de mayor edad-más probabilidad de padecimiento y dietas basadas en considerables proteínas y grasas en conjunto a la falta de fibras y vegetales.

Justificación

Tenemos que tener en claro la importancia del porqué de este tipo de investigaciones centradas en tipos de cánceres específicos y el uso de herramientas computacionales para su solución y desarrollo.

Según el IMSS en México “se diagnostican cerca de 15 mil casos nuevos de personas con cáncer de colon y/o recto, enfermedad que al tratarse de manera oportuna, permite la curación en nueve de cada 10 casos” [7] , sin embargo, alrededor de 6 mil 500 personas mueren al año por este padecimiento, es por este tipo de datos que la recopilación de información sobre pacientes es muy importante, pues esto nos podría permitir hacer un diagnóstico más temprano de la enfermedad, pero esto no sería posible de hacer sin las herramientas informáticas que nos facilita la biología computacional con la cual podemos incluso ver cosas tan esenciales para la documentación como aquellos genes diferencialmente expresados, lo cual es justamente lo que buscamos dentro de este proyecto, el poder describir cómo la ciencias de la computación contribuyen en las ciencias de la salud para la mejora de la vida de la población mexicana y del mundo, pues cuando se hace uso de la tecnología en la medicina se pueden obtener información de manera más eficaz y asertiva, siendo capaces de identificar las amenazas a tiempo.

Objetivo

En este proyecto propondremos un análisis de datos a través del uso herramientas computacionales sobre la información genética en pacientes que padezcan de cáncer de colon y de aquellos en condiciones normales para que podamos realizar comparaciones e identifiquemos la información necesaria para la detección temprana de este tipo de cáncer en pacientes jóvenes con la enfermedad.

Métodos

Para poder identificar las diferencias de medias usando la base de datos “Multi_Cancer_Data.RData”, se observó que tenía varias muestras de distintos tipos de tumor y de tejido normal en diferentes órganos, clasificadas por genes. Calculamos cuántas muestras eran de cada tipo, buscando cuántas columnas tenían la palabra tumor, así se identificaron 190 tumores y 90 normales. Dividimos las muestras en dos variables con ayuda de la función `grep`, una que contenga los tumores y otra los normales, con esto podemos calcular las medias por gen de cada uno, así como calcular la diferencia entre estas. Organizamos estos valores en un data frame y lo ordenamos de manera que podamos identificar aquellos genes en los que se tenga mayor diferencia entre las muestras de tumores y las de normales.

Después repetimos el proceso, pero esta vez buscando muestras de tumor colorrectal y del normal del colon, se calcula la media para cada gen, la diferencia, y se organiza en un data frame, el cual también se ordena de mayor a menor diferencia. Por último, buscamos si entre los primeros 20 genes se encuentran coincidencias con la comparación general, sin embargo, no se encontró ninguna.

Al analizar la base de datos “TCGA_COADREAD_comp_data.RData”, se encontró que se tenían dos objetos, uno con la expresión en genes, y otro con la clasificación de muestras entre jóvenes adultos y adultos mayores. Lo primero que se hizo fue separar las muestras usando la función `which`.

Posteriormente se crearon dos matrices vacías, en las cuales se pondrá la media por gen de las muestras en jóvenes y la media en adultos mayores; en la primera se pondrá la diferencia entre estas, mientras que en la segunda se añade el valor del p-value y fold change. A través de un ciclo `for` se irán calculando las medias y la diferencia, las cuales se concatenan para incluirlas en la primera matriz. De la misma forma se hace un ciclo para la otra matriz, el p-value se obtiene con un t-test entre las muestras de jóvenes y adultos mayores; dado que a los valores que tenemos ya se les ha aplicado el logaritmo, para el fold change solo se hace una resta; finalmente estos valores se meten a la matriz, y se crea un data frame con ella. Por último, la primera matriz se ordena de mayor a menor diferencia, mientras que la otra se ordena de menor a mayor p-value, dado que mientras más pequeño sea este, más significativa la diferencia que se presenta.

Una vez obtenida esta lista, por el p value más significativo, nos quedamos con los primeros 20 renglones. Eso nos da la información de 20 los genes con diferencia más significativa.

Para visualizar la expresión diferencial, haremos un heatmap con la función `heatmap.2` de `gplots` en R. Para ello previamente haremos distinción entre los genes que están sobreexpresados (Fold change de jóvenes respecto adultos mayores positivo) y los sub-expresados (Fold change negativo), después lo contrastaremos con las 286 muestras de los adultos mayores y los jóvenes.

La escala del FC irá desde -2 coloreando morado oscuro hasta 2 representado en naranja oscuro. En las columnas el naranja representa a las muestras de adultos mayores y el violeta las muestras jóvenes, mientras que en los renglones el naranja localiza a los genes sobreexpresados y el violeta a los sub-expresados. La representación podría no ser muy informativa al principio, por lo que tenemos que hacer un cambio de escala, normalizando los datos. Además, la jerarquización podría no producir una separación obvia entre las características, así que pasaremos a “refinar” la gráfica haciendo que todas las muestras de un tipo sean contiguas. Añadiremos líneas de separación para que el cambio sea más evidente.

Resultados y discusión

En la fig 3 pueden ver los 10 genes que tienen una mayor diferencia de medias de expresión genética entre muestras con tumor colorrectal y muestras de tejido normal en el colon, siendo así los más afectados por este tipo de cáncer. La comparación se hace así porque no tendría sentido

usar el tejido normal de otro 6rgano, de esa manera los datos no tendrían relevancia para el caso de estudio.

	Tumor_Colon	Normal_Colon	Diff
<i>Selenium binding protein 1_RC_AA290679_at</i>	0.129818181818182	3.53390909090909	3.40409090909091
<i>DRA Down-regulated in adenoma_L02785_at</i>	0.392090909090909	3.63054545454545	3.23845454545455
<i>HBG2 Hemoglobin gamma-G_M10050_at</i>	0.723636363636364	3.812	3.08836363636364
<i>clone 428431 3', mRNA sequence. (from Genbank)_RC_AA004415_at</i>	-0.175090909090909	2.91154545454545	3.08663636363636
<i>Vasoactive Intestinal Peptide_HG2987-HT3136_s_at</i>	-0.175636363636364	2.81818181818182	2.99381818181818
<i>IgG Fc binding protein_D84239_at</i>	0.00536363636363635	2.91018181818182	2.90481818181818
<i>IgG Fc binding protein_D84239_at-2</i>	0.00536363636363635	2.91018181818182	2.90481818181818
<i>Carcinoembryonic antigen family member 2_CGM2_X98311_at</i>	0.234	3.04681818181818	2.81281818181818
<i>Alcohol dehydrogenase 3 (class I), gamma polypeptide_M12272_s_at</i>	-0.333454545454545	2.416	2.74945454545455
<i>CA2 Carbonic anhydrase II_Y00339_s_at</i>	-0.155818181818182	2.50363636363636	2.65945454545455

Figura 3: Mayores diferencias de la expresión génica del cáncer del colon y el tejido normal

Por otro lado, en la figura 4 se encuentran los genes con las 20 diferencias de medias más significativas. En genes como el MTERF, EL TRMT12 o el CCDC67 la expresión media de los jóvenes es mayor a la de los adultos, y en genes como PIWIL1, ZNF239 Y ZNF600 es menor. Con ayuda de las intensidades de color en el mapa de calor podemos ubicar cuán diferente es la expresión de un gen, si fuese de un naranja oscuro el cambio se mostraría como muy expresado, si fuese morado oscuro estaría muy sub-expresado. De hecho con ayuda del heatmap vemos que efectivamente la subexpresión de las muestras jóvenes se concentra en los genes PIWIL1, ZNF239 y ZNF600, además del KCNRG. También es fácil ubicar el cuadro donde están los genes que se expresan más en los jóvenes que en los viejos, porque en la columna de los jóvenes predomina el naranja de ellos, y en la columna de los adultos mayores el morado.

	Young	Old	P value	Fold change
<i>MTERF</i>	8.39193345806447	7.62581493724862	8.05479940160979e-07	0.766118520815851
<i>PRND</i>	4.1636589588239	2.94486353109472	4.74801895022005e-06	1.21879542772918
<i>FZD9</i>	4.51446091911179	3.39788725291058	1.00073864409562e-05	1.11657366620122
<i>MLF1</i>	8.38633153278814	7.45551423446731	1.63055286117958e-05	0.930817298320829
<i>PCSK1N</i>	5.16845587203589	3.54927214981964	7.66185753359232e-05	1.61918372221625
<i>LOC100009676</i>	7.22734164279482	7.01135376406205	9.19970222044128e-05	0.215987878732773
<i>PIWIL1</i>	4.27838391057039	5.78701577664172	9.9453155794296e-05	-1.50863186607134
<i>KCNRG</i>	4.13745416279735	4.60794562186288	0.000172625065691156	-0.470491459065527
<i>ZNF239</i>	7.58108540379969	8.07940465360042	0.000185618894822931	-0.498319249800725
<i>KCNS3</i>	8.48503258677779	7.88636081615188	0.000189306870994985	0.598671770625909
<i>C5orf49</i>	1.52838523410566	0.946149249345996	0.000235664741202597	0.582235984759665
<i>CCDC90B</i>	10.3562861002724	10.1679942883161	0.00029335909374069	0.188291811956333
<i>TRMT12</i>	9.10249051271302	8.7368067603745	0.000301590546892821	0.365683752338523
<i>GAL</i>	7.78198621175521	6.64964532123932	0.00033632201097631	1.13234089051589
<i>SPATA17</i>	4.79342438485741	4.20255351331262	0.000352336782370654	0.590870871544787
<i>GATA4</i>	3.3170109978803	1.59991197884909	0.000355111278291191	1.71709901903121
<i>ZNF75A</i>	8.0651152220732	7.66977095280017	0.000419408076008467	0.395344269273034
<i>GAMT</i>	6.18290849166571	5.59042002829649	0.000468213427466913	0.592488463369223
<i>CCDC67</i>	1.57932996641175	0.943719242049466	0.000572116789871022	0.635610724362281
<i>ZNF600</i>	8.93837335613552	9.19021469955713	0.000615068601570442	-0.251841343421606

Figura 4: Los 20 genes con diferencia más significativa entre muestras de adultos mayores y adultos jóvenes

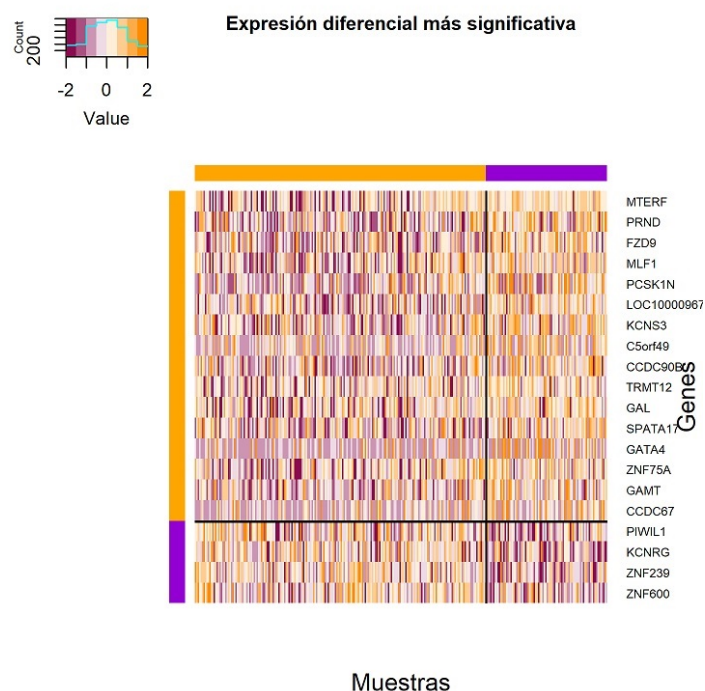


Figura 5: Heatmap con la información de los 20 genes con diferencia más significativa entre muestras de adultos mayores y adultos jóvenes

Durante una experimentación, se recolectan numerosos datos que tienen la posibilidad de variar en formato y provenir de distintos tipos de procesos. Por esta razón, la realización de muestreos significativos, el planteamiento de variables adecuadas y la selección de modelos estadísticos que mejor las representen, cobra gran relevancia si se desea incrementar la confiabilidad de un estudio.

Uno de estos modelos estadísticos es el valor p, el cual indica la probabilidad de que la hipótesis nula sea cierta. La hipótesis nula que cabe en nuestro análisis es que no se observará una diferencia entre la expresión de genes de jóvenes y la de personas de mayor edad.[8] Por ello, un p-value pequeño (menor a 5 %) indica que los individuos de la muestra fueron seleccionados de tal forma que es posible aceptar la hipótesis alternativa de que sí existe una diferencia significativa. En el presente análisis se calculó el p-value mediante una prueba t de Student para cada gen. En el caso de los genes donde se observaron las mayores diferencias, el p-value es lo suficientemente pequeño para confiar en los resultados, como se muestra a continuación.

	Young	Old	P.value	Fold.change
<i>GATA4</i>	3.3170109978803	1.59991197884909	0.000355111278291191	1.71709901903121
<i>PCSK1N</i>	5.16845587203589	3.54927214981964	7.66185753359232e-05	1.61918372221625
<i>XIST</i>	8.05519005336508	6.68569834204682	0.0620767822236754	1.36949171131826
<i>DUSP27</i>	7.19730307054406	5.87769535921682	0.0286972003946733	1.31960771132723
<i>HAVCR1</i>	4.64232643533573	3.33793596089615	0.00109583905799871	1.30439047443958
<i>DSC3</i>	7.69216054603897	6.40933196790677	0.00135584123091032	1.28282857813219
<i>DKK1</i>	4.81203714881848	3.55429781129549	0.000690245932001636	1.25773933752299
<i>PRND</i>	4.1636589588239	2.94486353109472	4.74801895022005e-06	1.21879542772918
<i>FOLR1</i>	5.9031030903056	4.73088880875393	0.00152428690934921	1.17221428155167
<i>CPS1</i>	6.13608370774431	4.97889312648362	0.006534873912239	1.15719058126068

Figura 6: Ejemplos de comparaciones de medias con p values significativos

A toda la manipulación de datos anterior se le puede dar una interpretación biológica útil para el desarrollo de tratamientos o para darle una explicación a las enfermedades. De los datos provenientes del conjunto Multi.Cancer.Data, fue posible observar que existe una menor expresión de genes relacionados con el colon en células cancerosas que en células sanas. Considerando definición de cáncer que dice que se caracteriza por el desarrollo de células anormales, se puede decir que estas células cuyo ADN contiene genes menormente expresados posiblemente sean la causa de que las células se desarrollen anormalmente, ya que las funciones de la célula se ven afectadas comparadas con el wildtype del código genético.

Esta observación nos lleva a los resultados del análisis que se hizo de los datos en TCGA_COADREAD, donde encontramos que en aquellos genes más diferencialmente expresados se observa una menor expresión en adultos mayores que en jóvenes. Con esto último podemos decir que la probabilidad de contraer cáncer incrementa con la edad. Eso está altamente relacionado con la tendencia que observamos en los gráficos de los datos de México, donde se tienen más personas de edad avanzada con cáncer que jóvenes.

Conclusión

El desarrollo de este análisis no se lleva a la conclusión de que es pertinente hacer un dispositivo capaz de detectar la presencia de genes relacionados con el colon que estén menormente expresados para determinar si la persona que realiza la prueba es más proclive a contraer cáncer en el colon. Además, es importante tomar en consideración la edad de la persona para realizar recomendaciones más adecuadas de en qué momento se debe hacer esta prueba.

El uso de tecnologías computacionales acelera muchos procesos de análisis biológico facilitando así pruebas médicas importantes como en el caso de la detección del cáncer de colon. La constante evolución de estas tecnologías nos permite ahorrar recursos económicos y tiempo, para dar una respuesta más rápida a las necesidades de los pacientes y mejorando el bienestar social.

Conclusiones personales

Ana Ivette Farías Rodríguez

Las actividades vistas durante estas 5 semanas, fueron de gran utilidad al momento de resolver la situación problema, incluso podríamos decir como la situación problema es un recopilación de estas actividades, ejemplo de esto son las actividades y las cuales fueron nuestra base para poder resolver la problemática de la materia gracias a su gran parecido en cuanto a lo que se debe hacer. Respecto a las competencias, en la competencia SICT0200 la mayoría de la actividades se trataron sobre manejar bases de datos para con éstas identificar patrones o coincidencias; la SEG0500 fue una de las más trabajadas pues en cada una de las actividades se puso a prueba nuestro razonamiento lógico para poder llegar a la resolución de las tarea, finalmente competencias como SICT0400 y SEG0600 se trabajaron dentro de los trabajos de la materia, como por ejemplo la comunicación, pues se tenía que dar a entender lo que se estaba escribiendo en el código y fuera de este, fue cuando se tenían que escribir reportes adicionales y finalmente las base de todos los trabajos era aplicar herramientas computacionales para el beneficio de la población.

José Gerardo Villanueva Barceló

Los temas y actividades vistos en clase fueron pilar para la solución de la situación problema, se nos propiciaron las herramientas, soportes y métodos para realizar los trabajos de manera correcta, se nos enseñó a la búsqueda de la información de bases de datos respectivas de los temas, aplicación de metodos en un lenguaje nuevo como R studio y el análisis de los datos generados para llegar a una conclusión. Respecto a las competencias el resultado fue crear un documento que cumpla con los estándares de divulgación, aplicación de herramientas para el análisis de los datos e investigación de fuentes óptimas, todo esto producto de las actividades vistas en esta

materia. Consideró increíble el impacto que tiene el cáncer colorrectal en el mundo y México, la importancia de analizar estos padecimiento es sumamente importante ya que pueden presentar riesgo de mortalidad en la población, en lo particular pienso que todos los países necesitan análisis preventivos. En cuanto a R studio considero que es un lenguaje completo y óptimo para el análisis de datos, como soy de una carrera en tecnologías computacionales, ayudó a la comprensión del correcto manejo de datos, filtros y presentación de resultados mediante gráficas.

Luis Humberto Sánchez Vaca

Las actividades ayudaron bastante para la situación planteada, ya que gracias a los análisis que fuimos haciendo a lo largo de ellas pudimos obtener más información sobre el cáncer de colon, así como también aprendimos a refinar la búsqueda de artículos útiles para lo que estamos investigando. También en la actividad 1 aprendimos a manejar bases de datos para encontrar información relevante sobre genes relacionados con cáncer colorrectal, entramos a NCBI y obtuvimos datos como localización, longitud, tejido donde más se expresan, funciones, etc. En cuanto a las competencias, se trabajó en casi todas las actividades con modelos computacionales en RStudio para el análisis de datos buscando encontrar resultados relevantes para la situación problema, junto con esto hubo actividades bastante complejas que requirieron de mucha búsqueda para encontrar las herramientas adecuadas para su solución. Además, con el proyecto es importante que la información se comunique de manera precisa y entendible, así cómo usar fuentes confiables como las que se nos proveyó con las bases de datos; esto con el objetivo de ayudar a la población a tener un poco más de información sobre el cáncer de colon.

José de Jesús Gutiérrez Aldrete

Considero que esta ha sido una de las unidades de formación mejor planeadas que he tenido hasta ahora. Las actividades estaban muy bien hiladas y tenían propósito. Fue enriquecedor conocer las diferencias entre un tejido normal de colon y un tumor en el colon, sin embargo, desde luego que iba a haber diferencias. Lo que me brindó de un mayor panorama fue en la actividad 4 saber de las diferencias de expresión génica de las personas que tuvieron cáncer de colon a temprana edad contra las personas mayores de 50, jesa diferencia no la habría imaginado! Con razón es un fenómeno tan estudiado y ha tenido tal relevancia para ser nuestra situación problema. Fue grato conocer herramientas de probabilidad y estadística a través de la bioinformática, ahora quiero hacer un mapa de calor para clasificar a mis compañeros de carrera. Finalmente, ha sido un buen reto abstraer el funcionamiento de R, en suma de conocer las sintaxis y relaciones de matlab, python y C++. Los datos pueden darnos la respuesta a muchas cosas, y a través de esta secuencia he trabajado en manipularlos, analizarlos y comunicarlos. Soy de ciencia de datos y la U.F. me ha ayudado a consolidar la importancia de estos para el bienestar de la sociedad.

Óscar Miranda Escalante

R y RStudio son herramientas computacionales (entre muchas otras) que me permitieron analizar grandes cantidades de datos provenientes de diversas fuentes para poder construir modelos que representaran adecuadamente esta información. Lo anterior me llevó a aplicar aprendizajes del área de computación y tecnologías de la información a una investigación que involucra conceptos científicos (biología y medicina). Esto me dio la oportunidad de relacionar las áreas que parecen ajenas a mi avenida y explorar alternativas en el uso de herramientas computacionales que permitan economizar procesos para incrementar la sustentabilidad de las aplicaciones digitales en términos de consumo energético y de tiempo. A su vez, me he dado cuenta del potencial que tienen herramientas de programación como software para el desarrollo e implementación de soluciones que pueden incluso salvar vidas.

Referencias

- [1] Grados, O. (2014, julio 19). ¿Qué es el colon y cuáles son sus funciones? Consultado el 29 de abril de 2020. <http://www.medicarteoncologia.com/que-es-el-colon-y-cuales-son-sus-funciones/>
- [2] Puente, J. y Velasco, G. (2019, diciembre 16). ¿Qué es el cáncer y cómo se desarrolla?. <https://seom.org/informacion-sobre-el-cancer/que-es-el-cancer-y-como-se-desarrolla>
- [3] NIH. (2015, febrero 9). ¿Qué es el cáncer?. <https://www.cancer.gov/espanol/cancer/naturaleza/que-es>
- [4] CDC. (2020, febrero 10). Cuáles son los factores de riesgo del cáncer colorrectal?. https://www.cdc.gov/spanish/cancer/colorectal/basic_info/risk_factors.htm
- [5] INCAN. (2017, junio 19). Datos y recursos Pacientes de Nuevo Ingreso. <https://datos.gob.mx/busca/dataset/pacientes-de-nuevo-ingreso>
- [6] Cervantes, G. y Ochoa, F. (2008, noviembre 20). Epidemiología del cáncer de colon y recto. <http://www.gamomeo.com/temp/SUPLEMENTO%20V7%20No%204%202008%20CANCER%20DE%20COLON.pdf>
- [7] IMSS. (2019, marzo). En México cada año se diagnostican cerca de 15 mil nuevos casos de de personas con cáncer de colon. <http://www.imss.gob.mx/prensa/archivo/201903/070>
- [8] Molina, M. (2017). ¿Qué significa realmente el valor de p?. Scientific Electronic Library Online. Consultado el 01 de mayo. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1139-76322017000500014

Apéndice: casos de cáncer colorrectal en la población mexicana, INCAN, 2016

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
1	6	86	Masculino	DISTRITO FEDERAL	TLALPAN	TUMOR MALIGNO DEL RECTO
2	43	51	Femenino	GUERRERO	CHILPANCINGO DE LOS BRAVO	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
3	80	58	Femenino	PUEBLA	PUEBLA	TUMOR MALIGNO DEL RECTO
4	85	76	Masculino	MORELOS	XOCHITEPEC	TUMOR MALIGNO DEL COLON
5	108	72	Masculino	HIDALGO	ACATLAN	TUMOR MALIGNO DEL COLON
6	132	49	Femenino	MEXICO	CHALCO	TUMOR MALIGNO DEL COLON
7	140	68	Femenino	DISTRITO FEDERAL	TLALPAN	TUMOR MALIGNO DEL COLON
8	145	54	Masculino	MEXICO	TULTEPEC	TUMOR MALIGNO DEL COLON
9	152	82	Femenino	HIDALGO	IXMIQUILPAN	TUMOR MALIGNO DEL RECTO
10	156	36	Masculino	MEXICO	HUIXQUILUCAN	TUMOR MALIGNO DEL RECTO
11	170	58	Femenino	DISTRITO FEDERAL	AZCAPOTZALCO	TUMOR MALIGNO DEL RECTO
12	180	35	Masculino	DISTRITO FEDERAL	XOCHIMILCO	TUMOR MALIGNO DEL COLON
13	187	89	Femenino	MEXICO	VALLE DE CHALCO SOLIDARIDAD	TUMOR MALIGNO DEL RECTO
14	219	80	Masculino	MEXICO	VILLA DEL CARBON	TUMOR MALIGNO DEL COLON
15	232	60	Masculino	MEXICO	ECATEPEC DE MORELOS	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
16	261	50	Masculino	DISTRITO FEDERAL	AZCAPOTZALCO	TUMOR MALIGNO DEL RECTO
17	271	85	Femenino	DISTRITO FEDERAL	XOCHIMILCO	TUMOR MALIGNO DEL RECTO

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
18	285	71	Masculino	DISTRITO FEDERAL	VENUSTIANO CARRANZA	TUMOR MALIGNO DEL COLON
19	302	81	Femenino	VERACRUZ	NARANJOS AMATLÁN	TUMOR MALIGNO DEL RECTO
20	324	75	Masculino	DISTRITO FEDERAL	BENITO JUÁREZ	TUMOR MALIGNO DEL RECTO
21	336	64	Femenino	PUEBLA	PAHUATLÁN	TUMOR MALIGNO DEL COLON
22	337	55	Femenino	MEXICO	CUAUTITLAN IZCALLI	TUMOR MALIGNO DEL COLON
23	365	51	Femenino	DISTRITO FEDERAL	MIGUEL HIDALGO	TUMOR MALIGNO DEL COLON
24	367	51	Masculino	DISTRITO FEDERAL	CUAUHTEMOC	TUMOR MALIGNO DEL COLON
25	382	39	Masculino	MORELOS	YAUTEPEC	TUMOR MALIGNO DEL RECTO
26	387	75	Femenino	MEXICO	ATLACOMULCO	TUMOR MALIGNO DEL COLON
27	400	25	Masculino	GUERRERO	CHILPANCINGO DE LOS BRAVO	TUMOR MALIGNO DEL RECTO
28	404	55	Masculino	VERACRUZ	ATZALAN	TUMOR MALIGNO DEL RECTO
29	453	61	Masculino	MEXICO	TOLUCA	TUMOR MALIGNO DEL RECTO
30	457	60	Femenino	MEXICO	TLALNEPANTLA DE BAZ	TUMOR MALIGNO DEL RECTO
31	461	59	Masculino	CHIAPAS	TUXTLA GUTIÉRREZ	TUMOR MALIGNO DEL RECTO
32	466	55	Femenino	DISTRITO FEDERAL	VENUSTIANO CARRANZA	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
33	478	59	Femenino	TLAXCALA	CHIAUTEMPAN	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
34	522	44	Masculino	OAXACA	SAN MIGUEL SOYALTEPEC	TUMOR MALIGNO DEL RECTO
35	524	39	Femenino	DISTRITO FEDERAL	XOCHIMILCO	TUMOR MALIGNO DEL COLON
36	549	51	Masculino	MEXICO	ATIZAPAN DE ZARAGOZA	TUMOR MALIGNO DEL COLON
37	569	47	Masculino	DISTRITO FEDERAL	XOCHIMILCO	TUMOR MALIGNO DEL RECTO
38	598	62	Masculino	MORELOS	TETELA DEL VOLCÁN	TUMOR MALIGNO DEL RECTO
39	608	65	Masculino	GUERRERO	ACAPULCO DE JUÁREZ	TUMOR MALIGNO DEL RECTO
40	651	72	Masculino	MEXICO	ACAMBAY	TUMOR MALIGNO DEL COLON
41	663	66	Femenino	PUEBLA	ZACAPOXTLA	TUMOR MALIGNO DEL COLON
42	668	54	Femenino	VERACRUZ	COSOLEACAQUE	TUMOR MALIGNO DEL COLON
43	679	49	Femenino	DISTRITO FEDERAL	VENUSTIANO CARRANZA	TUMOR MALIGNO DEL COLON
44	735	37	Masculino	MEXICO	TEJUPILCO	TUMOR MALIGNO DEL RECTO
45	780	44	Masculino	CHIAPAS	VENUSTIANO CARRANZA	TUMOR MALIGNO DEL RECTO
46	797	53	Masculino	MEXICO	CUAUTITLAN IZCALLI	TUMOR MALIGNO DEL RECTO
47	810	31	Femenino	DISTRITO FEDERAL	TLALPAN	TUMOR MALIGNO DEL COLON
48	921	73	Femenino	VERACRUZ	SAN JUAN EVANGELISTA	TUMOR MALIGNO DEL RECTO
49	922	35	Masculino	MORELOS	CUERNAVACA	TUMOR MALIGNO DEL COLON

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
50	928	70	Femenino	DISTRITO FEDERAL	BENITO JUÁREZ	TUMOR MALIGNO DEL COLON
51	944	19	Masculino	COAHUILA	SALTILLO	TUMOR MALIGNO DEL COLON
52	948	74	Femenino	MEXICO	ECATEPEC DE MORELOS	TUMOR MALIGNO DEL COLON
53	949	41	Masculino	MEXICO	ZINACANTEPEC	TUMOR MALIGNO DEL RECTO
54	964	75	Masculino	DISTRITO FEDERAL	TLALPAN	TUMOR MALIGNO DEL COLON
55	1026	67	Femenino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL RECTO
56	1041	53	Femenino	OAXACA	SAN SEBASTIÁN TECOMAXTLAHUACA	TUMOR MALIGNO DEL RECTO
57	1044	35	Femenino	MEXICO	CHICONCUAC	TUMOR MALIGNO DEL RECTO
58	1063	70	Masculino	DISTRITO FEDERAL	BENITO JUÁREZ	TUMOR MALIGNO DEL RECTO
59	1193	69	Masculino	DISTRITO FEDERAL	BENITO JUÁREZ	TUMOR MALIGNO DEL COLON
60	1208	56	Femenino	SINALOA	MAZATLÁN	TUMOR MALIGNO DEL RECTO
61	1216	46	Masculino	DISTRITO FEDERAL	VENUSTIANO CARRANZA	TUMOR MALIGNO DEL COLON
62	1231	83	Masculino	BAJA CALIFORNIA	MEXICALI	TUMOR MALIGNO DEL COLON
63	1270	60	Femenino	VERACRUZ	ILAMATLÁN	TUMOR MALIGNO DEL COLON
64	1277	45	Masculino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL COLON
65	1279	25	Femenino	MORELOS	TEMOAC	TUMOR MALIGNO DEL RECTO

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
66	1282	60	Masculino	DISTRITO FEDERAL	VENUSTIANO CARRANZA	TUMOR MALIGNO DEL COLON
67	1298	46	Masculino	DISTRITO FEDERAL	CUAUHTEMOC	TUMOR MALIGNO DEL COLON
68	1304	56	Masculino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL RECTO
69	1320	66	Masculino	HIDALGO	TULANCINGO DE BRAVO	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
70	1338	67	Masculino	DISTRITO FEDERAL	COYOACAN	TUMOR MALIGNO DEL COLON
71	1355	49	Masculino	MORELOS	ZACATEPEC	TUMOR MALIGNO DEL COLON
72	1358	73	Masculino	DISTRITO FEDERAL	CUAUHTEMOC	TUMOR MALIGNO DEL COLON
73	1366	29	Femenino	DISTRITO FEDERAL	AZCAPOTZALCO	TUMOR MALIGNO DEL COLON
74	1369	59	Femenino	DISTRITO FEDERAL	CUAUHTEMOC	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
75	1387	39	Femenino	HIDALGO	TULANCINGO DE BRAVO	TUMOR MALIGNO DEL RECTO
76	1399	57	Femenino	MEXICO	ATLACOMULCO	TUMOR MALIGNO DEL COLON
77	1402	49	Masculino	DISTRITO FEDERAL	COYOACAN	TUMOR MALIGNO DEL COLON
78	1404	47	Masculino	OAXACA	SAN MIGUEL SANTA FLOR	TUMOR MALIGNO DEL COLON
79	1414	66	Masculino	DISTRITO FEDERAL	COYOACAN	TUMOR MALIGNO DEL COLON
80	1422	58	Femenino	DISTRITO FEDERAL	BENITO JUÁREZ	TUMOR MALIGNO DEL RECTO
81	1424	70	Masculino	MEXICO	SAN MATEO ATENCO	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
82	1431	38	Femenino	MEXICO	CHIMALHUACAN	TUMOR MALIGNO DEL RECTO
83	1434	62	Masculino	DISTRITO FEDERAL	XOCHIMILCO	TUMOR MALIGNO DEL COLON
84	1435	60	Masculino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL RECTO
85	1446	46	Masculino	MEXICO	NICOLAS ROMERO	TUMOR MALIGNO DEL COLON
86	1458	84	Femenino	MORELOS	HUITZILAC	TUMOR MALIGNO DEL COLON
87	1463	52	Masculino	DISTRITO FEDERAL	ALVARO OBREGÓN	TUMOR MALIGNO DEL COLON
88	1488	56	Femenino	DISTRITO FEDERAL	TLALPAN	TUMOR MALIGNO DEL RECTO
89	1522	66	Masculino	MEXICO	TULTITLAN	TUMOR MALIGNO DEL RECTO
90	1526	33	Masculino	PUEBLA	ZACATLAN	TUMOR MALIGNO DEL RECTO
91	1585	81	Femenino	GUERRERO	TEOLOAPAN	TUMOR MALIGNO DEL COLON
92	1589	53	Masculino	DISTRITO FEDERAL	ALVARO OBREGÓN	TUMOR MALIGNO DEL COLON
93	1592	60	Femenino	HIDALGO	ZAPOTLÁN DE JUAREZ	TUMOR MALIGNO DEL RECTO
94	1595	53	Femenino	MEXICO	CHIMALHUACAN	TUMOR MALIGNO DEL RECTO
95	1625	79	Masculino	DISTRITO FEDERAL	ALVARO OBREGÓN	TUMOR MALIGNO DEL COLON
96	1638	47	Femenino	MEXICO	LA PAZ	TUMOR MALIGNO DEL RECTO
97	1652	74	Femenino	DISTRITO FEDERAL	AZCAPOTZALCO	TUMOR MALIGNO DEL COLON

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
98	1668	71	Femenino	MEXICO	NAUCALPAN	TUMOR MALIGNO DEL COLON
99	1687	45	Masculino	MEXICO	NICOLAS ROMERO	TUMOR MALIGNO DEL RECTO
100	1730	60	Femenino	DISTRITO FEDERAL	AZCAPOTZALCO	TUMOR MALIGNO DEL RECTO
101	1734	63	Femenino	DISTRITO FEDERAL	CUAUHTEMOC	TUMOR MALIGNO DEL COLON
102	1737	31	Masculino	MEXICO	JUCHITEPEC	TUMOR MALIGNO DEL RECTO
103	1746	37	Masculino	MEXICO	ECATEPEC	TUMOR MALIGNO DEL RECTO
104	1749	52	Femenino	MICHOACAN	VILLAMAR	TUMOR MALIGNO DEL COLON
105	1760	64	Femenino	MEXICO	TLALNEPANTLA DE BAZ	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
106	1809	80	Masculino	DISTRITO FEDERAL	AZCAPOTZALCO	TUMOR MALIGNO DEL RECTO
107	1845	64	Femenino	DISTRITO FEDERAL	ALVARO OBREGÓN	TUMOR MALIGNO DEL COLON
108	1877	64	Masculino	HIDALGO	EL ARENAL	TUMOR MALIGNO DEL RECTO
109	1881	86	Masculino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL COLON
110	1914	65	Masculino	HIDALGO	PACHUCA DE SOTO	TUMOR MALIGNO DEL RECTO
111	1940	63	Femenino	DISTRITO FEDERAL	IZTACALCO	TUMOR MALIGNO DEL RECTO
112	1941	77	Femenino	DISTRITO FEDERAL	XOCHIMILCO	TUMOR MALIGNO DEL RECTO
113	1956	54	Femenino	MORELOS	TEMIXCO	TUMOR MALIGNO DEL COLON

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
114	1960	62	Masculino	PUEBLA	TEHUACAN	TUMOR MALIGNO DEL COLON
115	1972	71	Masculino	HIDALGO	ALMOLOYA	TUMOR MALIGNO DEL RECTO
116	1974	66	Masculino	MEXICO	NEZAHUALCOYOTL	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
117	1988	59	Masculino	HIDALGO	NOPALA DE VILLAGRAN	TUMOR MALIGNO DEL RECTO
118	1991	47	Masculino	MEXICO	CALIMAYA	TUMOR MALIGNO DEL RECTO
119	2063	51	Masculino	MEXICO	TLALNEPANTLA DE BAZ	TUMOR MALIGNO DEL RECTO
120	2078	61	Masculino	HIDALGO	ACTOPAN	TUMOR MALIGNO DEL COLON
121	2083	56	Femenino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL COLON
122	2097	34	Femenino	SINALOA	MOCORITO	TUMOR MALIGNO DEL COLON
123	2150	43	Femenino	MEXICO	ZINACANTEPEC	TUMOR MALIGNO DEL COLON
124	2152	65	Femenino	HIDALGO	IXMIQUILPAN	TUMOR MALIGNO DEL COLON
125	2194	49	Femenino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL RECTO
126	2197	53	Masculino	MEXICO	NAUCALPAN DE JUAREZ	TUMOR MALIGNO DEL RECTO
127	2200	46	Masculino	PUEBLA	PUEBLA	TUMOR MALIGNO DEL COLON
128	2206	55	Femenino	DISTRITO FEDERAL	MILPA ALTA	TUMOR MALIGNO DEL RECTO
129	2223	60	Femenino	DISTRITO FEDERAL	CUAJIMALPA DE MORELOS	TUMOR MALIGNO DEL RECTO

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
130	2239	64	Femenino	CHIAPAS	COMITAN DE DOMINGUEZ	TUMOR MALIGNO DEL COLON
131	2243	66	Masculino	DISTRITO FEDERAL	BENITO JUÁREZ	TUMOR MALIGNO DEL RECTO
132	2255	36	Femenino	DISTRITO FEDERAL	MILPA ALTA	TUMOR MALIGNO DEL RECTO
133	2256	50	Femenino	MEXICO	TECAMAC	TUMOR MALIGNO DEL RECTO
134	2296	63	Masculino	MORELOS	EMILIANO ZAPATA	TUMOR MALIGNO DEL RECTO
135	2299	34	Masculino	HIDALGO	METEPEC	TUMOR MALIGNO DEL COLON
136	2313	66	Masculino	PUEBLA	ZACATLAN	TUMOR MALIGNO DEL RECTO
137	2315	52	Masculino	DISTRITO FEDERAL	GUSTAVO A MADERO	TUMOR MALIGNO DEL RECTO
138	2347	30	Masculino	DISTRITO FEDERAL	ALVARO OBREGÓN	TUMOR MALIGNO DEL COLON
139	2364	71	Femenino	JALISCO	ZAPOPAN	TUMOR MALIGNO DEL RECTO
140	2410	48	Femenino	DISTRITO FEDERAL	GUSTAVO A MADERO	TUMOR MALIGNO DEL RECTO
141	2426	68	Masculino	MEXICO	NEZAHUALCOYOTL	TUMOR MALIGNO DEL RECTO
142	2429	74	Femenino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL RECTO
143	2457	57	Femenino	MEXICO	TENANCINGO	TUMOR MALIGNO DEL COLON
144	2463	72	Femenino	DISTRITO FEDERAL	IZTAPALAPA	TUMOR MALIGNO DEL COLON
145	2469	45	Femenino	MICHOACAN	SUSUPUATO	TUMOR MALIGNO DEL COLON

	FOLIO	EDAD	SEXO	ESTADO	MUNICIPIO	DESCRIPCION.DIAGNOSTICO
146	2517	54	Masculino	DISTRITO FEDERAL	TLAHUAC	TUMOR MALIGNO DEL RECTO
147	2528	72	Femenino	MEXICO	SAN MATEO ATENCO	TUMOR MALIGNO DEL RECTO
148	2534	65	Femenino	HIDALGO	TIZAYUCA	TUMOR MALIGNO DEL COLON
149	2536	68	Femenino	DISTRITO FEDERAL	TLAHUAC	TUMOR MALIGNO DEL COLON
150	2537	22	Masculino	MICHOACAN	JUNGAPEO	TUMOR MALIGNO DEL COLON
151	2587	54	Femenino	DISTRITO FEDERAL	VENUSTIANO CARRANZA	TUMOR MALIGNO DEL COLON
152	2595	67	Femenino	MEXICO	TEXCOCO	TUMOR MALIGNO DEL COLON
153	2603	43	Masculino	PUEBLA	CHIGNAHUAPAN	TUMOR MALIGNO DEL COLON
154	2607	56	Masculino	HIDALGO	TIZAYUCA	TUMOR MALIGNO DEL RECTO
155	2638	58	Masculino	MEXICO	TLALNEPANTLA DE BAZ	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA
156	2646	33	Masculino	MEXICO	CAPULHUAC	TUMOR MALIGNO DEL COLON
157	2711	77	Masculino	MEXICO	ACULCO	TUMOR MALIGNO DEL COLON
158	2758	73	Masculino	DISTRITO FEDERAL	ALVARO OBREGÓN	TUMOR MALIGNO DEL COLON
159	2782	46	Masculino	GUERRERO	CHILPANCINGO DE LOS BRAVO	TUMOR MALIGNO DEL RECTO
160	2791	34	Femenino	MORELOS	TEMOAC	TUMOR MALIGNO DEL RECTO
161	2810	49	Femenino	DISTRITO FEDERAL	BENITO JUÁREZ	TUMOR MALIGNO DE LA UNION RECTOSIGMOIDEA