Reporte Tecnico

Fernando Chávez 142353, Bernardo Mondragón 143743, Rayan García 144424

17 de octubre de 2016

## 1. Introducción

El Proyecto consiste en el análisis de la incidencia en mordeduras de perro en México del año 2004 a 2015. Dicha incidencia se ordena por Fuente de Notificación, Mes de Ocurrencia y Grupo de Edad. Los datos usados en este proyecto provienen de la secretaría de salud: <http://www.epidemiologia.salud.gob.mx/anuario/html/anuarios.html>.

El objetivo es presentar una serie de gráficas y análisis de cómo ha cambiado la incidencia de las mordeduras de perro a través de los años. Creemos firmemente que el gran objetivo del análisis de datos en el sector salud es prevenir.

Es importante presentar de manera clara y organizada los datos de las distintas clasificaciones (i.e fuente de notificación, mese de ocurrencia y grupo de edad), como menciona Edward Tufte en su lectura de Beautiful Evidence "las presentaciones deben ser intelectuales, que puedan transmitir conocimiento, que enseñen y que digan algo. Porque el problema con muchas presentaciones hoy en día es que tienen mucho contenido pero no dicen nada, no transmiten una bella evidencia".

También creemos que los datos y la información deben seguir con los principios que Stephen Few menciona en su lectura Statistical Narrative. Es decir, esta debe ser simple, informativa, verdadera, contextual, concreta y procesable. También no nos olvidamos de agrupar la información de manera correcta y mostrar la documentación y las comparaciones de su análisis, como lo menciona Edward Tufte.

Creemos que el análisis de datos es importante para el sector salud, ya que podría utilizarse para conocer la tendencia, la frecuencia y la variabilidad de los datos. Ya que si el sector salud conoce de manera detallada la incidencia de las mordeduras de perro en México puede tomar mejores decisiones para atender a la población. Se pueden crear campañas o se pueden enfocar más en los estados "problemáticos".

En este trabajo se pudieron incluir más variables a analizar, sin embargo esto hubiese implicado perder de vista el objetivo principal del curso, el cual es poder presentar un análisis de datos desde el principio, esto es, desde la captura, la limpieza y la organización de datos.

## 2. Definición del Problema

En la obtención de los datos surgen varias preguntas a resolver y aclarar: ¿Cuál es la tendencia de las mordeduras de perro a través de los años?, ¿Cómo se cambió la incidencia de las mordeduras de perro en los estados cada año?, ¿Cómo se fueron reportando las mordeduras de perro en las fuentes de notificación a través de los años?, ¿Cómo se reportaron las mordeduras en los grupos de edad a través de los años?, ¿Cuáles son lo estados que están por arriba de la media en cada año?, ¿Cuáles son los máximos y mínimos a través de los años?, ¿Cómo es la tendencia de cada estado en cada año?

Para poder resolver nuestras preguntas tuvimos que plantaer nuestro análisis en cinco divisiones: La primera sería conocer el máximo, el mínimo, la media, la varianza, la desviación estandar, y los cuartiles, esto es para conocer mejor la probabilidad y la organización de nuestros datos en el tiempo.

La segunda pregunta surge de un pensamiento racional y clásico: ¿Cómo se están comportando las mordeduras en México a través de los años? es decir la tendencia.

La tercera, sería analizar cada una de nuestras variables(Fuente de Notificación, Mes de Ocurrencia y Grupo de Edad) incluyendo también por cada uno de nuestros estados el cambio en el tiempo en una multigráfica para ver con mayor presición como los años cambian las mordeduras en el tiempo.

La cuarta,consideramos relevante el uso del tipo de gráfica Boxplot en nuestro análisis, ya que con esto podemos analizar estadísticamente mejor el acumulado de las probabilidades y ver con mayor presición la mediana, la media, y lo cuartiles, al igual que los valores atípicos que registra cada gráfica. Para esto consideramos primordiales cuatro boxplots, incluyendo en el eje de las ordenadas el Total de mordeduras y en el eje de las abscisas los años, comprandolos con: Estados, meses, grupo de edad y fuente de notificación.

Para la quinta, es necesario presentar gráficas en forma de pirámide para ver la comparación de los datos entre mujers y hombres.

## 3. Obtención y Descripción de los Datos

### 3.1 Descarga de Archivos

Se descargaron los PDF's de la pagina consultada para cada año desde 2004 a 2015, tanto para Fuente de Notificación, Mes de Ocurrencia, y Grupo de Edad incluyendo: General, Hombres y Mujeres, 108 en total

Después para poderlos leer en Rstudio, primero se convirtió a formato Excel utilizando la página de internet: <https://online2pdf.com/en/pdf2excel>.

Al momento de leerlos en Excel se observó lo siguiente:

#### 3.1.1 Fuente de Notificación

En cada archivo excel de Fuente de Notificación, se encontraba en cada columna: "Estado", Salud", "Imss.ord", "Issste", "Imss.op", "Dif", "Pemex", "Sedena", "Semar", "Otras" y "Total. Siendo "OTRAS", otras fuentes notificación diferentes a las anteriores y Total al total obtenido de cada renglón. También se obtuvo en el excel un renglón de Totales y encabezados de la descripción de cada archivo.

#### 3.1.2 Mes de Ocurrencia

En cada excel de Mes de Ocurrencia, se encontraba en cada columna: "Estado" "ENE", "FEB", "MAR", "ABR", "MAY", "JUN", "JUL", "AGO", "SEP", "OCT", "NOV", "DIC". Dependiendo del año que se descargara el PDF había en algunos una columna con la tasa utilizada para calcular los resultados. También se podían hallar los meses sin abreviaturas y con la primera letra mayúscula, al igual que en algunos la columna de Total variaba en la primera columna o en la última.

Para eso decidimos pertinente dejar todas las columas con las abreviaturas de los meses, eliminar la columna de tasa en todos archivos excel y los espacios encontrados entre columnas, ya que a veces al transformar de PDF a excel puede generar columnas vacías.

#### 3.1.3 Grupo de Edad

En cada archivo excel de Grupo de Edad, se encuentra en cada columna: "Estado", ">1", "1-4", "5-9", "10-14", "15-19", "20-24", "25-44", "45-49", "50-59", "60-64", ">65", "Ign".

Uno de los principales problemas, que se consideraron de esta variable, es que los intervalos de las edades no son constantes, esto es, que hay intervalos de cuatro años al igual que hay un intervalo de 19 años el cual presentará mayor probabilidad de que los datos estén contenidos en este. Esto impide que se pueda hacer un mejor análisis ya que los intervalos no son los mismos. En estadística uno de los principios básicos para poder hacer un análisis es que poder comparar con las mismas medidas ya que en otro caso se pueden distorsionar los resultados, como fue en este caso.

Otro de los problemas que presenta cada excel de Grupo de Edad es que su segunda columna ">1" se puede interpretar como mayor a uno pero con la tercera columna "1-4" se puede crear confusión al momento de interpretar los datos.Probablemente lo que quizo mencionar el archivo es que ">1" indica menor a un año.

#### 3.1.4 Similaridades en todos los Formatos

En la columna Estados se encontraban 32 en total, cada uno con Mayúscula al principio y con su respectivo acento, en caso que tuviera. Se indica a la Ciudad de México como "Distrito Federal" y al Estado de México como "México".

En algunas diapositivas se encuentran "N.A." y "S/R" indicando ambas que no hubo registro de datos en ese apartado.

En cada excel, independientemente de cada Fuente, se encontraba en ciertas ocaciones columnas vacías, datos comprimidos en una misma celda, renglón de Totales y leyendas de la información de cada archivo. En ciertos casos nos vimos obligados a modificar los archivos porque para programar algunos cambios en Rstudio hubieran sido muy complicado, especialmente se tuvo problemas con todos los archivos del año 2012 ya que eliminaba la columna de Estados y la colocaba en otro espacio al igual que los datos los modificaba.

Fueron pocos en los que nos vimos obligados a cambiar a mano la información obtenida para el archivo excel, ya que en la conversión de formato se había modificado considerablemente.

### 3.2 Formato CSV

Para poder leer los archivos de excel en Rstudio, se tuvieron que guardar en el formato CSV, ya guardados se pudieron leer sin ningún problema en Rstudio.

## 4. Calidad de los Datos

En cuanto a la calidad de los datos, se tuvo que hacer pruebas de consistencia, esto consiste en probar que el total de los archivos de GENERAL fueran iguales a HOMBRES más MUJERES. Se creo un código para comprobar la igualdad en Rstudio, en caso de que no se diera la igualdad se arrojaba el año en el cual había habido inconsistencia en caso contrario no arrojaba nada. A final de cuentas se comprobó que no había ninguna inconsistencia en la igualdad de ningún año. También se comprobó que los totales fueran correctos en cada Formato, Fuente de Notificación, Mes de Ocurrencia y Grupo de Edad.

### 4.1 Prueba de Consistencia de Fuente de Notificación

## [[1]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2004 SALUD 103044 57823 45221 103044 0  
## 2 2004 IMSS.ORD 14379 8648 5731 14379 0  
## 3 2004 ISSSTE 3192 1686 1506 3192 0  
## 4 2004 IMSS.OP 8843 4550 4293 8843 0  
## 5 2004 DIF 423 217 206 423 0  
## 6 2004 PEMEX 696 402 294 696 0  
## 7 2004 SEDENA 431 265 166 431 0  
## 8 2004 SEMAR 155 91 64 155 0  
## 9 2004 OTRAS 5400 3215 2185 5400 0  
##   
## [[2]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2005 SALUD 100693 56292 44401 100693 0  
## 2 2005 IMSS.ORD 12751 7557 5194 12751 0  
## 3 2005 ISSSTE 3073 1637 1436 3073 0  
## 4 2005 IMSS.OP 9659 5129 4530 9659 0  
## 5 2005 DIF 290 160 130 290 0  
## 6 2005 PEMEX 724 390 334 724 0  
## 7 2005 SEDENA 337 208 129 337 0  
## 8 2005 SEMAR 122 71 51 122 0  
## 9 2005 OTRAS 5483 3085 2398 5483 0  
##   
## [[3]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2006 SALUD 99492 55505 43987 99492 0  
## 2 2006 IMSS.ORD 14164 8411 5753 14164 0  
## 3 2006 ISSSTE 2961 1576 1385 2961 0  
## 4 2006 IMSS.OP 10743 5567 5176 10743 0  
## 5 2006 DIF 247 128 119 247 0  
## 6 2006 PEMEX 774 419 355 774 0  
## 7 2006 SEDENA 392 248 144 392 0  
## 8 2006 SEMAR 88 45 43 88 0  
## 9 2006 OTRAS 4747 2743 2004 4747 0  
##   
## [[4]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2007 SALUD 87374 48848 38526 87374 0  
## 2 2007 IMSS.ORD 11537 6783 4754 11537 0  
## 3 2007 ISSSTE 2653 1334 1319 2653 0  
## 4 2007 IMSS.OP 9285 4947 4338 9285 5  
## 5 2007 DIF 315 169 146 315 12  
## 6 2007 PEMEX 802 445 357 802 18  
## 7 2007 SEDENA 307 201 106 307 1  
## 8 2007 SEMAR 106 58 48 106 14  
## 9 2007 OTRAS 4453 2508 1945 4453 5  
##   
## [[5]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2008 SALUD 84250 46895 37355 84250 0  
## 2 2008 IMSS.ORD 12862 7333 5529 12862 0  
## 3 2008 ISSSTE 2949 1540 1409 2949 0  
## 4 2008 IMSS.OP 8670 4572 4098 8670 5  
## 5 2008 DIF 352 199 153 352 13  
## 6 2008 PEMEX 993 517 476 993 18  
## 7 2008 SEDENA 232 163 69 232 0  
## 8 2008 SEMAR 108 65 43 108 14  
## 9 2008 OTRAS 4213 2426 1787 4213 6  
##   
## [[6]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2009 SALUD 75897 41763 34134 75897 0  
## 2 2009 IMSS.ORD 12612 7298 5314 12612 0  
## 3 2009 ISSSTE 2616 1374 1242 2616 0  
## 4 2009 IMSS.OP 9121 4916 4205 9121 5  
## 5 2009 DIF 254 103 151 254 13  
## 6 2009 PEMEX 795 400 395 795 18  
## 7 2009 SEDENA 255 156 99 255 0  
## 8 2009 SEMAR 136 84 52 136 14  
## 9 2009 OTRAS 4255 2405 1850 4255 6  
##   
## [[7]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2010 SALUD 71902 39660 32242 71902 0  
## 2 2010 IMSS.ORD 13267 7603 5664 13267 0  
## 3 2010 ISSSTE 3021 1542 1479 3021 0  
## 4 2010 IMSS.OP 8944 4785 4159 8944 5  
## 5 2010 DIF 318 181 137 318 13  
## 6 2010 PEMEX 851 463 388 851 18  
## 7 2010 SEDENA 277 160 117 277 0  
## 8 2010 SEMAR 171 117 54 171 14  
## 9 2010 OTRAS 4557 2610 1947 4557 4  
##   
## [[8]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2011 SALUD 74352 41404 32948 74352 0  
## 2 2011 IMSS.ORD 15721 9019 6702 15721 0  
## 3 2011 ISSSTE 2987 1470 1517 2987 0  
## 4 2011 IMSS.OP 10066 5383 4683 10066 5  
## 5 2011 DIF 463 252 211 463 13  
## 6 2011 PEMEX 1001 517 484 1001 18  
## 7 2011 SEDENA 292 187 105 292 0  
## 8 2011 SEMAR 156 85 71 156 14  
## 9 2011 OTRAS 5056 2798 2258 5056 3  
##   
## [[9]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2012 SALUD 79985 44374 35611 79985 0  
## 2 2012 IMSS.ORD 16859 9652 7207 16859 0  
## 3 2012 ISSSTE 2985 1529 1456 2985 0  
## 4 2012 IMSS.OP 11061 5810 5251 11061 5  
## 5 2012 DIF 464 241 223 464 13  
## 6 2012 PEMEX 1036 566 470 1036 18  
## 7 2012 SEDENA 322 210 112 322 0  
## 8 2012 SEMAR 122 60 62 122 13  
## 9 2012 OTRAS 5377 3006 2371 5377 3  
##   
## [[10]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2013 SALUD 78100 43689 34411 78100 0  
## 2 2013 IMSS.ORD 6271 3538 2733 6271 0  
## 3 2013 ISSSTE 3427 1768 1659 3427 0  
## 4 2013 IMSS.OP 10199 5367 4832 10199 5  
## 5 2013 DIF 414 224 190 414 13  
## 6 2013 PEMEX 997 497 500 997 18  
## 7 2013 SEDENA 309 186 123 309 1  
## 8 2013 SEMAR 127 73 54 127 13  
## 9 2013 OTRAS 5120 2870 2250 5120 3  
##   
## [[11]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2014 SALUD 78830 43543 35287 78830 0  
## 2 2014 IMSS.ORD 19332 10959 8373 19332 0  
## 3 2014 ISSSTE 3779 1919 1860 3779 0  
## 4 2014 IMSS.OP 8988 4661 4327 8988 5  
## 5 2014 DIF 445 248 197 445 13  
## 6 2014 PEMEX 1019 555 464 1019 18  
## 7 2014 SEDENA 300 181 119 300 1  
## 8 2014 SEMAR 163 90 73 163 13  
## 9 2014 OTRAS 5197 2938 2259 5197 3  
##   
## [[12]]  
## ANIO FUENTE MORD.G MORD.H MORD.M H + M NA's  
## 1 2015 SALUD 73169 40025 33144 73169 0  
## 2 2015 IMSS.ORD 19859 11070 8789 19859 0  
## 3 2015 ISSSTE 3410 1679 1731 3410 0  
## 4 2015 IMSS.OP 7696 4018 3678 7696 5  
## 5 2015 DIF 525 279 246 525 13  
## 6 2015 PEMEX 1009 558 451 1009 18  
## 7 2015 SEDENA 293 178 115 293 1  
## 8 2015 SEMAR 156 88 68 156 13  
## 9 2015 OTRAS 5412 3045 2367 5412 3

### 4.2 Prueba de Consistencia para Mes de Ocurrencia

## [[1]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2004 ENE 9089 5139 3950 9089 0  
## 2 2004 FEB 10052 5668 4384 10052 0  
## 3 2004 MAR 15317 8626 6691 15317 0  
## 4 2004 ABR 11351 6391 4960 11351 0  
## 5 2004 MAY 11766 6738 5028 11766 0  
## 6 2004 JUN 14664 8246 6418 14664 0  
## 7 2004 JUL 11120 6368 4752 11120 0  
## 8 2004 AGO 13691 7825 5866 13691 0  
## 9 2004 SEP 10434 5741 4693 10434 0  
## 10 2004 OCT 10559 5829 4730 10559 0  
## 11 2004 NOV 11382 6318 5064 11382 0  
## 12 2004 DIC 7138 4008 3130 7138 0  
##   
## [[2]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2005 ENE 9272 5226 4046 9272 0  
## 2 2005 FEB 10130 5622 4508 10130 0  
## 3 2005 MAR 13611 7618 5993 13611 0  
## 4 2005 ABR 11448 6346 5102 11448 0  
## 5 2005 MAY 14234 7959 6275 14234 0  
## 6 2005 JUN 11482 6364 5118 11482 0  
## 7 2005 JUL 10862 6274 4588 10862 0  
## 8 2005 AGO 13056 7329 5727 13056 0  
## 9 2005 SEP 10409 5800 4609 10409 0  
## 10 2005 OCT 9971 5548 4423 9971 0  
## 11 2005 NOV 11118 6184 4934 11118 0  
## 12 2005 DIC 7539 4259 3280 7539 0  
##   
## [[3]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2006 ENE 11405 6415 4990 11405 0  
## 2 2006 FEB 10249 5726 4523 10249 0  
## 3 2006 MAR 11358 6353 5005 11358 0  
## 4 2006 ABR 11287 6387 4900 11287 0  
## 5 2006 MAY 14972 8345 6627 14972 0  
## 6 2006 JUN 11822 6551 5271 11822 0  
## 7 2006 JUL 10581 5973 4608 10581 0  
## 8 2006 AGO 13297 7504 5793 13297 0  
## 9 2006 SEP 10731 5965 4766 10731 0  
## 10 2006 OCT 12309 6841 5468 12309 0  
## 11 2006 NOV 8510 4623 3887 8510 0  
## 12 2006 DIC 7087 3959 3128 7087 0  
##   
## [[4]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2007 ENE 9789 5428 4361 9789 0  
## 2 2007 FEB 9414 5336 4078 9414 0  
## 3 2007 MAR 10171 5655 4516 10171 0  
## 4 2007 ABR 11985 6780 5205 11985 0  
## 5 2007 MAY 10056 5618 4438 10056 0  
## 6 2007 JUN 10237 5663 4574 10237 0  
## 7 2007 JUL 12079 6864 5215 12079 0  
## 8 2007 AGO 9257 5221 4036 9257 0  
## 9 2007 SEP 9379 5098 4281 9379 0  
## 10 2007 OCT 10279 5709 4570 10279 0  
## 11 2007 NOV 7535 4149 3386 7535 0  
## 12 2007 DIC 6651 3772 2879 6651 0  
##   
## [[5]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2008 ENE 7600 4251 3349 7600 0  
## 2 2008 FEB 8643 4786 3857 8643 0  
## 3 2008 MAR 9012 5031 3981 9012 0  
## 4 2008 ABR 13389 7540 5849 13389 0  
## 5 2008 MAY 9973 5425 4548 9973 0  
## 6 2008 JUN 9630 5367 4263 9630 0  
## 7 2008 JUL 11654 6564 5090 11654 0  
## 8 2008 AGO 9023 5098 3925 9023 0  
## 9 2008 SEP 10548 5847 4701 10548 0  
## 10 2008 OCT 8398 4549 3849 8398 0  
## 11 2008 NOV 7437 4053 3384 7437 0  
## 12 2008 DIC 9322 5199 4123 9322 0  
##   
## [[6]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2009 ENE 8036 4552 3484 8036 0  
## 2 2009 FEB 8801 4903 3898 8801 0  
## 3 2009 MAR 12100 6610 5490 12100 0  
## 4 2009 ABR 9055 5092 3963 9055 0  
## 5 2009 MAY 9154 5006 4148 9154 0  
## 6 2009 JUN 11267 6145 5122 11267 0  
## 7 2009 JUL 8742 4924 3818 8742 0  
## 8 2009 AGO 8439 4655 3784 8439 0  
## 9 2009 SEP 9549 5202 4347 9549 0  
## 10 2009 OCT 7339 4008 3331 7339 0  
## 11 2009 NOV 6680 3607 3073 6680 0  
## 12 2009 DIC 6779 3795 2984 6779 0  
##   
## [[7]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2010 ENE 7041 3918 3123 7041 0  
## 2 2010 FEB 7557 4258 3299 7557 0  
## 3 2010 MAR 10779 5948 4831 10779 0  
## 4 2010 ABR 8985 4976 4009 8985 0  
## 5 2010 MAY 9125 5030 4095 9125 0  
## 6 2010 JUN 11305 6211 5094 11305 0  
## 7 2010 JUL 8603 4720 3883 8603 0  
## 8 2010 AGO 8715 4915 3800 8715 0  
## 9 2010 SEP 9938 5474 4464 9938 0  
## 10 2010 OCT 7692 4206 3486 7692 0  
## 11 2010 NOV 6495 3557 2938 6495 0  
## 12 2010 DIC 7073 3908 3165 7073 0  
##   
## [[8]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2011 ENE 7138 4070 3068 7138 0  
## 2 2011 FEB 7858 4282 3576 7858 0  
## 3 2011 MAR 11655 6415 5240 11655 0  
## 4 2011 ABR 8993 4972 4021 8993 0  
## 5 2011 MAY 9124 5080 4044 9124 0  
## 6 2011 JUN 11495 6267 5228 11495 0  
## 7 2011 JUL 9513 5361 4152 9513 0  
## 8 2011 AGO 11232 6282 4950 11232 0  
## 9 2011 SEP 8598 4730 3868 8598 0  
## 10 2011 OCT 8175 4533 3642 8175 0  
## 11 2011 NOV 9569 5350 4219 9569 0  
## 12 2011 DIC 6744 3773 2971 6744 0  
##   
## [[9]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2012 ENE 10422 5957 4465 10422 0  
## 2 2012 FEB 9149 5074 4075 9149 0  
## 3 2012 MAR 9827 5401 4426 9827 0  
## 4 2012 ABR 9709 5393 4316 9709 0  
## 5 2012 MAY 12643 6836 5807 12643 0  
## 6 2012 JUN 10407 5722 4685 10407 0  
## 7 2012 JUL 12439 6923 5516 12439 0  
## 8 2012 AGO 9632 5384 4248 9632 0  
## 9 2012 SEP 9467 5260 4207 9467 0  
## 10 2012 OCT 11174 6208 4966 11174 0  
## 11 2012 NOV 7423 4038 3385 7423 0  
## 12 2012 DIC 5919 3252 2667 5919 0  
##   
## [[10]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2013 ENE 8807 4918 3889 8807 0  
## 2 2013 FEB 8031 4489 3542 8031 0  
## 3 2013 MAR 7784 4386 3398 7784 0  
## 4 2013 ABR 8749 4852 3897 8749 0  
## 5 2013 MAY 11267 6239 5028 11267 0  
## 6 2013 JUN 9318 5121 4197 9318 0  
## 7 2013 JUL 11231 6277 4954 11231 0  
## 8 2013 AGO 8505 4836 3669 8505 0  
## 9 2013 SEP 8216 4553 3663 8216 0  
## 10 2013 OCT 9644 5317 4327 9644 0  
## 11 2013 NOV 7061 3732 3329 7061 0  
## 12 2013 DIC 6351 3492 2859 6351 0  
##   
## [[11]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2014 ENE 9550 5291 4259 9550 0  
## 2 2014 FEB 8890 4879 4011 8890 0  
## 3 2014 MAR 9467 5197 4270 9467 0  
## 4 2014 ABR 11562 6359 5203 11562 0  
## 5 2014 MAY 10160 5646 4514 10160 0  
## 6 2014 JUN 10352 5732 4620 10352 0  
## 7 2014 JUL 12271 6840 5431 12271 0  
## 8 2014 AGO 9349 5213 4136 9349 0  
## 9 2014 SEP 8904 4867 4037 8904 0  
## 10 2014 OCT 11288 6213 5075 11288 0  
## 11 2014 NOV 8007 4282 3725 8007 0  
## 12 2014 DIC 8253 4575 3678 8253 0  
##   
## [[12]]  
## ANIO MES MORD.G MORD.H MORD.M H + M NA's  
## 1 2015 ENE 7470 4197 3273 7470 0  
## 2 2015 FEB 8377 4591 3786 8377 0  
## 3 2015 MAR 11480 6214 5266 11480 0  
## 4 2015 ABR 9759 5300 4459 9759 0  
## 5 2015 MAY 9448 5104 4344 9448 0  
## 6 2015 JUN 11939 6424 5515 11939 0  
## 7 2015 JUL 9202 4992 4210 9202 0  
## 8 2015 AGO 11218 6233 4985 11218 0  
## 9 2015 SEP 8636 4786 3850 8636 0  
## 10 2015 OCT 8301 4503 3798 8301 0  
## 11 2015 NOV 9528 5177 4351 9528 0  
## 12 2015 DIC 6171 3419 2752 6171 0

### 4.3 Prueba de Consistencia para Grupo de Edad

## [[1]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2004 MAYOR.A.UNO 487 282 205 487 0  
## 2 2004 UNO.A.CUATRO 13801 8573 5228 13801 0  
## 3 2004 CINCO.A.NUEVE 24796 15298 9498 24796 0  
## 4 2004 DIEZ.A.CATORCE 21325 13816 7509 21325 0  
## 5 2004 QUINCE.A.DIECINUEVE 12087 6734 5353 12087 0  
## 6 2004 VEINTE.A.VEINTICUATRO 9801 5323 4478 9801 0  
## 7 2004 VEINTICINCO.A.CUARENTAYCUATRO 27616 13626 13990 27616 0  
## 8 2004 CUARENTAYCINCO.A.CUARENTAYNUEVE 5550 2562 2988 5550 0  
## 9 2004 CINCUENTA.A.CINCUENTAYNUEVE 8498 4028 4470 8498 0  
## 10 2004 SESENTA.A.SESENTAYCUATRO 4040 2022 2018 4040 0  
## 11 2004 MAYOR.A.SESENTAYCINCO 8480 4576 3904 8480 0  
## 12 2004 NO.SE.REPORTO.LA.EDAD 82 57 25 82 0  
##   
## [[2]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2005 MAYOR.A.UNO 361 218 143 361 0  
## 2 2005 UNO.A.CUATRO 13351 8124 5227 13351 0  
## 3 2005 CINCO.A.NUEVE 23925 15099 8826 23925 0  
## 4 2005 DIEZ.A.CATORCE 20594 13298 7296 20594 0  
## 5 2005 QUINCE.A.DIECINUEVE 11793 6623 5170 11793 0  
## 6 2005 VEINTE.A.VEINTICUATRO 9063 4903 4160 9063 0  
## 7 2005 VEINTICINCO.A.CUARENTAYCUATRO 26990 13083 13907 26990 0  
## 8 2005 CUARENTAYCINCO.A.CUARENTAYNUEVE 5598 2459 3139 5598 0  
## 9 2005 CINCUENTA.A.CINCUENTAYNUEVE 8676 4008 4668 8676 0  
## 10 2005 SESENTA.A.SESENTAYCUATRO 3919 1976 1943 3919 0  
## 11 2005 MAYOR.A.SESENTAYCINCO 8751 4674 4077 8751 0  
## 12 2005 NO.SE.REPORTO.LA.EDAD 111 64 47 111 0  
##   
## [[3]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2006 MAYOR.A.UNO 404 274 130 404 0  
## 2 2006 UNO.A.CUATRO 13089 8027 5062 13089 0  
## 3 2006 CINCO.A.NUEVE 23948 14970 8978 23948 0  
## 4 2006 DIEZ.A.CATORCE 19830 12849 6981 19830 0  
## 5 2006 QUINCE.A.DIECINUEVE 11762 6607 5155 11762 0  
## 6 2006 VEINTE.A.VEINTICUATRO 9135 4946 4189 9135 0  
## 7 2006 VEINTICINCO.A.CUARENTAYCUATRO 27313 13298 14015 27313 0  
## 8 2006 CUARENTAYCINCO.A.CUARENTAYNUEVE 5776 2586 3190 5776 0  
## 9 2006 CINCUENTA.A.CINCUENTAYNUEVE 9070 4170 4900 9070 0  
## 10 2006 SESENTA.A.SESENTAYCUATRO 3918 1988 1930 3918 0  
## 11 2006 MAYOR.A.SESENTAYCINCO 9286 4891 4395 9286 0  
## 12 2006 NO.SE.REPORTO.LA.EDAD 77 36 41 77 0  
##   
## [[4]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2007 MAYOR.A.UNO 240 143 97 240 0  
## 2 2007 UNO.A.CUATRO 11358 7105 4253 11358 0  
## 3 2007 CINCO.A.NUEVE 20638 12909 7729 20638 0  
## 4 2007 DIEZ.A.CATORCE 17037 11134 5903 17037 0  
## 5 2007 QUINCE.A.DIECINUEVE 10623 5956 4667 10623 0  
## 6 2007 VEINTE.A.VEINTICUATRO 7944 4247 3697 7944 0  
## 7 2007 VEINTICINCO.A.CUARENTAYCUATRO 23894 11702 12192 23894 0  
## 8 2007 CUARENTAYCINCO.A.CUARENTAYNUEVE 5054 2224 2830 5054 0  
## 9 2007 CINCUENTA.A.CINCUENTAYNUEVE 8188 3736 4452 8188 0  
## 10 2007 SESENTA.A.SESENTAYCUATRO 3489 1705 1784 3489 0  
## 11 2007 MAYOR.A.SESENTAYCINCO 8291 4392 3899 8291 0  
## 12 2007 NO.SE.REPORTO.LA.EDAD 76 40 36 76 0  
##   
## [[5]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2008 MAYOR.A.UNO 275 159 116 275 0  
## 2 2008 UNO.A.CUATRO 11128 6836 4292 11128 0  
## 3 2008 CINCO.A.NUEVE 20268 12606 7662 20268 0  
## 4 2008 DIEZ.A.CATORCE 15976 10402 5574 15976 0  
## 5 2008 QUINCE.A.DIECINUEVE 10421 5918 4503 10421 0  
## 6 2008 VEINTE.A.VEINTICUATRO 7970 4384 3586 7970 0  
## 7 2008 VEINTICINCO.A.CUARENTAYCUATRO 23420 11400 12020 23420 0  
## 8 2008 CUARENTAYCINCO.A.CUARENTAYNUEVE 5296 2332 2964 5296 0  
## 9 2008 CINCUENTA.A.CINCUENTAYNUEVE 7952 3626 4326 7952 0  
## 10 2008 SESENTA.A.SESENTAYCUATRO 3514 1727 1787 3514 0  
## 11 2008 MAYOR.A.SESENTAYCINCO 8334 4275 4059 8334 0  
## 12 2008 NO.SE.REPORTO.LA.EDAD 75 45 30 75 0  
##   
## [[6]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2009 MAYOR.A.UNO 246 146 100 246 0  
## 2 2009 UNO.A.CUATRO 10234 6263 3971 10234 0  
## 3 2009 CINCO.A.NUEVE 18753 11663 7090 18753 0  
## 4 2009 DIEZ.A.CATORCE 14424 9224 5200 14424 0  
## 5 2009 QUINCE.A.DIECINUEVE 9635 5406 4229 9635 0  
## 6 2009 VEINTE.A.VEINTICUATRO 7340 3941 3399 7340 0  
## 7 2009 VEINTICINCO.A.CUARENTAYCUATRO 21929 10777 11152 21929 0  
## 8 2009 CUARENTAYCINCO.A.CUARENTAYNUEVE 4854 2111 2743 4854 0  
## 9 2009 CINCUENTA.A.CINCUENTAYNUEVE 7674 3436 4238 7674 0  
## 10 2009 SESENTA.A.SESENTAYCUATRO 3215 1611 1604 3215 0  
## 11 2009 MAYOR.A.SESENTAYCINCO 7463 3898 3565 7463 0  
## 12 2009 NO.SE.REPORTO.LA.EDAD 174 23 151 174 0  
##   
## [[7]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2010 MAYOR.A.UNO 395 284 111 395 0  
## 2 2010 UNO.A.CUATRO 9940 6106 3834 9940 0  
## 3 2010 CINCO.A.NUEVE 17888 11079 6809 17888 0  
## 4 2010 DIEZ.A.CATORCE 14625 9430 5195 14625 0  
## 5 2010 QUINCE.A.DIECINUEVE 9414 5323 4091 9414 0  
## 6 2010 VEINTE.A.VEINTICUATRO 7292 3952 3340 7292 0  
## 7 2010 VEINTICINCO.A.CUARENTAYCUATRO 20976 10199 10777 20976 0  
## 8 2010 CUARENTAYCINCO.A.CUARENTAYNUEVE 4749 2126 2623 4749 0  
## 9 2010 CINCUENTA.A.CINCUENTAYNUEVE 7479 3297 4182 7479 0  
## 10 2010 SESENTA.A.SESENTAYCUATRO 3198 1558 1640 3198 0  
## 11 2010 MAYOR.A.SESENTAYCINCO 7301 3738 3563 7301 0  
## 12 2010 NO.SE.REPORTO.LA.EDAD 51 29 22 51 0  
##   
## [[8]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2011 MAYOR.A.UNO 261 155 106 261 0  
## 2 2011 UNO.A.CUATRO 10821 6561 4260 10821 0  
## 3 2011 CINCO.A.NUEVE 18718 11672 7046 18718 0  
## 4 2011 DIEZ.A.CATORCE 15609 10158 5451 15609 0  
## 5 2011 QUINCE.A.DIECINUEVE 9706 5543 4163 9706 0  
## 6 2011 VEINTE.A.VEINTICUATRO 7812 4263 3549 7812 0  
## 7 2011 VEINTICINCO.A.CUARENTAYCUATRO 22447 11123 11324 22447 0  
## 8 2011 CUARENTAYCINCO.A.CUARENTAYNUEVE 5106 2263 2843 5106 0  
## 9 2011 CINCUENTA.A.CINCUENTAYNUEVE 8401 3709 4692 8401 0  
## 10 2011 SESENTA.A.SESENTAYCUATRO 3413 1617 1796 3413 0  
## 11 2011 MAYOR.A.SESENTAYCINCO 7745 4024 3721 7745 0  
## 12 2011 NO.SE.REPORTO.LA.EDAD 55 27 28 55 0  
##   
## [[9]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2012 MAYOR.A.UNO 276 150 126 276 0  
## 2 2012 UNO.A.CUATRO 11460 6917 4543 11460 0  
## 3 2012 CINCO.A.NUEVE 19775 12290 7485 19775 0  
## 4 2012 DIEZ.A.CATORCE 16501 10792 5709 16501 0  
## 5 2012 QUINCE.A.DIECINUEVE 10448 5991 4457 10448 0  
## 6 2012 VEINTE.A.VEINTICUATRO 8644 4785 3859 8644 0  
## 7 2012 VEINTICINCO.A.CUARENTAYCUATRO 24607 12156 12451 24607 0  
## 8 2012 CUARENTAYCINCO.A.CUARENTAYNUEVE 5666 2474 3192 5666 0  
## 9 2012 CINCUENTA.A.CINCUENTAYNUEVE 8934 3955 4979 8934 0  
## 10 2012 SESENTA.A.SESENTAYCUATRO 3605 1680 1925 3605 0  
## 11 2012 MAYOR.A.SESENTAYCINCO 8192 4196 3996 8192 0  
## 12 2012 NO.SE.REPORTO.LA.EDAD 103 62 41 103 0  
##   
## [[10]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2013 MAYOR.A.UNO 242 137 105 242 0  
## 2 2013 UNO.A.CUATRO 9665 5852 3813 9665 0  
## 3 2013 CINCO.A.NUEVE 16976 10579 6397 16976 0  
## 4 2013 DIEZ.A.CATORCE 14520 9594 4926 14520 0  
## 5 2013 QUINCE.A.DIECINUEVE 9588 5637 3951 9588 0  
## 6 2013 VEINTE.A.VEINTICUATRO 7592 4149 3443 7592 0  
## 7 2013 VEINTICINCO.A.CUARENTAYCUATRO 22168 10958 11210 22168 0  
## 8 2013 CUARENTAYCINCO.A.CUARENTAYNUEVE 5318 2336 2982 5318 0  
## 9 2013 CINCUENTA.A.CINCUENTAYNUEVE 8188 3604 4584 8188 0  
## 10 2013 SESENTA.A.SESENTAYCUATRO 3130 1473 1657 3130 0  
## 11 2013 MAYOR.A.SESENTAYCINCO 7408 3809 3599 7408 0  
## 12 2013 NO.SE.REPORTO.LA.EDAD 169 84 85 169 0  
##   
## [[11]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2014 MAYOR.A.UNO 283 158 125 283 0  
## 2 2014 UNO.A.CUATRO 10284 6143 4141 10284 0  
## 3 2014 CINCO.A.NUEVE 17997 11204 6793 17997 0  
## 4 2014 DIEZ.A.CATORCE 15779 10238 5541 15779 0  
## 5 2014 QUINCE.A.DIECINUEVE 10513 6022 4491 10513 0  
## 6 2014 VEINTE.A.VEINTICUATRO 9066 5098 3968 9066 0  
## 7 2014 VEINTICINCO.A.CUARENTAYCUATRO 25772 12921 12851 25772 0  
## 8 2014 CUARENTAYCINCO.A.CUARENTAYNUEVE 6048 2742 3306 6048 0  
## 9 2014 CINCUENTA.A.CINCUENTAYNUEVE 9580 4266 5314 9580 0  
## 10 2014 SESENTA.A.SESENTAYCUATRO 3946 1878 2068 3946 0  
## 11 2014 MAYOR.A.SESENTAYCINCO 8540 4384 4156 8540 0  
## 12 2014 NO.SE.REPORTO.LA.EDAD 245 40 205 245 0  
##   
## [[12]]  
## ANIO EDAD MORD.G MORD.H MORD.M H + M NA's  
## 1 2015 MAYOR.A.UNO 245 144 101 245 0  
## 2 2015 UNO.A.CUATRO 9153 5559 3594 9153 0  
## 3 2015 CINCO.A.NUEVE 16249 9972 6277 16249 0  
## 4 2015 DIEZ.A.CATORCE 13876 9038 4838 13876 0  
## 5 2015 QUINCE.A.DIECINUEVE 10235 5925 4310 10235 0  
## 6 2015 VEINTE.A.VEINTICUATRO 8753 4788 3965 8753 0  
## 7 2015 VEINTICINCO.A.CUARENTAYCUATRO 25065 12392 12673 25065 0  
## 8 2015 CUARENTAYCINCO.A.CUARENTAYNUEVE 5946 2712 3234 5946 0  
## 9 2015 CINCUENTA.A.CINCUENTAYNUEVE 9589 4271 5318 9589 0  
## 10 2015 SESENTA.A.SESENTAYCUATRO 3897 1818 2079 3897 0  
## 11 2015 MAYOR.A.SESENTAYCINCO 8449 4274 4175 8449 0  
## 12 2015 NO.SE.REPORTO.LA.EDAD 72 47 25 72 0

## 5. Transformación de los Datos

### 5.1 Limpieza

Se limpiaron y acomodaron los datos en Rstudio de tal forma que pudieran estar en un formato limpio y organizado, como lo mencionan Edward Tufte y Stephen Few en sus respectivas lecturas. No teniendo columnas innecesarias, datos extra y sin tener una información desorganizada.

La limpieza consistió en: eliminar espacios entre números, convertir todos los números en formato "numeric", cambiar todas las variables en mayúsculas, eliminar los espacios entre los nombres de los estados, quitar los acentos de todos los estados y por último cambiar "N.A." y "S/R" a NA's para que con Rstudio pudieramos hacer los análisis correspondientes.

En especial se modificó el Dataframe de Grupo de Edad ya que como Rstudio no lee variables como números tuvimos que cambiar las columnas originales a letra (i.e "1-4" se modificó a "UNO.A.CUATRO")

### 5.2 Organización

Después de hacer la limpieza para cada archivo se acomodó la información en formato organizado. Se realizó la lectura de los archivos, se añadieron para cada archivo un columna con el nombre de ANIO, indicando el año respectivo, se escogió el nombre de "ANIO" y no "AÑO" ya que en computadoras mac no se pueden leer "ñ's" en las variables. Además se agregó la columna de "SEXO", indicando a que archivo correspondian los datos si a: "GENERAL", "HOMBRES" o "MUJERES".

Después, se juntaron todas las tablas en una, para cada uno de los formatos, creando tres Dataframes, esto es, un Dataframe para Fuente de Notificación, Mes de Ocurrencia y Grupo de Edad, cada uno con las columnas: "ESTADO", "ANIO", Variable("FUENTE", "MES", "EDAD"), "SEXO" y Total de Mordeduras llamado "MORDEDURAS".

Además de las columnas de "ESTADO", "ANIO", "SEXO" y "MORDEDURAS" en la Variable de cada Dataframe está organizada de la siguiente manera:

#### 5.2.1 Fuente de Notificación

En el Dataframe de Fuente de Notificación, dentro de su variable "FUENTE", se encuentra: "SALUD", "IMSS.ORD", "ISSSTE", "IMSS.OP", "DIF", "PEMEX", "SEDENA", "SEMAR" y "OTRAS". Todas en mayúsculas y el espacio representado por un punto, si es necesario.

#### 5.2.2 Dataframe de Grupo de Edad

En el Dataframe de Grupo de Edad, dentro de la variable "EDAD", se encuentra: "MAYOR.A.UNO", "UNO.A.CUATRO", "CINCO.A.NUEVE", "DIEZ.A.CATORCE", "QUINCE.A.DIECINUEVE", "VEINTE.A.VEINTICUATRO", "VEINTICINCO.A.CUARENTAYCUATRO", "CUARENTAYCINCO.A.CUARENTAYNUEVE", "CINCUENTA.A.CINCUENTAYNUEVE", "SESENTA.A.SESENTAYCUATRO", "MAYOR.A.SESENTAYCINCO", "NO.SE.REPORTO.LA.EDAD". Cada dato presenta el intervalo de edad donde se presentaron mordeduras por perro todo en mayúsculas y un punto representando el espacio.

#### 5.2.3 Dataframe de Mes de Ocurrencia

En el Dataframe de Mes de Ocurrencia, dentro de la variable "MES", se encuentra: "ENE", "FEB", "MAR", "ABR", "MAY", "JUN", "JUL", "AGO", "SEP", "OCT", "NOV", "DIC". cada dato representado por la abreviatura de cada mes.

## 6. Análisis Exploratorio

### 6.1 Tablas con información básica

Para cada tabla se incluye el Año, La fuente con el valor máximo y su respectivo valor, la fuente con el valor mínimo y su respectivo valor, la media, la desvaición estandar, y cada cuartil.

#### 6.1.1 Información De Estados

## ANIO ESTADO.MAX MAX ESTADO.MIN MIN MEDIA VAR  
## 1 2004 Mexico 19556 Colima 624 4267.594 21863042  
## 2 2005 Mexico 18450 Baja California Sur 654 4160.375 19400789  
## 3 2006 Mexico 18972 Colima 506 4175.250 18029916  
## 4 2007 Mexico 15852 Colima 442 3651.000 12905316  
## 5 2008 Mexico 15555 Colima 517 3582.156 11754214  
## 6 2009 Mexico 13660 Colima 471 3310.656 9242410  
## 7 2010 Mexico 13538 Baja California Sur 551 3228.375 8333144  
## 8 2011 Mexico 14387 Colima 666 3440.438 9502038  
## 9 2012 Mexico 16686 Colima 686 3694.094 11718212  
## 10 2013 Mexico 16303 Colima 571 3280.125 11043588  
## 11 2014 Mexico 16972 Baja California Sur 787 3689.156 13093135  
## 12 2015 Mexico 15874 Baja California Sur 597 3485.281 11819072  
## STD 1st.Q 2nd.Q 3rd.Q 4rd.Q  
## 1 4675.793 624 1648.00 2646.0 4470.50  
## 2 4404.633 654 1704.25 2839.5 4572.75  
## 3 4246.165 506 1586.50 3028.0 4747.50  
## 4 3592.397 442 1400.50 2529.5 4164.50  
## 5 3428.442 517 1288.75 2250.0 4212.25  
## 6 3040.133 471 1149.25 2275.5 4092.75  
## 7 2886.718 551 1155.00 2440.0 3789.75  
## 8 3082.538 666 1253.00 2575.0 4295.00  
## 9 3423.187 686 1342.00 2705.5 4542.75  
## 10 3323.189 571 1258.25 2184.0 3924.25  
## 11 3618.444 787 1513.50 2591.0 4425.75  
## 12 3437.888 597 1604.25 2604.5 3941.00

Para la información de estados pudimos notar que en todos los años el máximo fue el Estado de México, y el mínimo fue entre Colima y Baja California Sur. La media fue oscilando entre 4300 y 3200 con uns desviación estandar muy grande esto se debe a los valores atípicos tan grandes que tenemos. Pero en promedio estaían mordiendo los perros como entre 4000 y 3000 personas.

#### 6.1.2 Información De Fuente de Notificación

## ANIO FUENTE.MAX MAX FUENTE.MIN MIN MEDIA VAR STD  
## 1 2004 SALUD 103044 SEMAR 155 15173.67 1108885098 33299.93  
## 2 2005 SALUD 100693 SEMAR 122 14792.44 1058232521 32530.49  
## 3 2006 SALUD 99492 SEMAR 88 14845.33 1032919592 32139.07  
## 4 2007 SALUD 87374 SEMAR 106 12981.33 795576233 28205.96  
## 5 2008 SALUD 84250 SEMAR 108 12736.56 738288154 27171.46  
## 6 2009 SALUD 75897 SEMAR 136 11771.22 597660410 24447.09  
## 7 2010 SALUD 71902 SEMAR 171 11478.67 533861490 23105.44  
## 8 2011 SALUD 74352 SEMAR 156 12232.67 570720272 23889.75  
## 9 2012 SALUD 79985 SEMAR 122 13134.56 661344049 25716.61  
## 10 2013 SALUD 78100 SEMAR 127 11662.67 632253830 25144.66  
## 11 2014 SALUD 78830 SEMAR 163 13117.00 645352584 25403.79  
## 12 2015 SALUD 73169 SEMAR 156 12392.11 558444104 23631.42  
## 1st.Q 2nd.Q 3rd.Q 4rd.Q  
## 1 155 431 3192 8843  
## 2 122 337 3073 9659  
## 3 88 392 2961 10743  
## 4 106 315 2653 9285  
## 5 108 352 2949 8670  
## 6 136 255 2616 9121  
## 7 171 318 3021 8944  
## 8 156 463 2987 10066  
## 9 122 464 2985 11061  
## 10 127 414 3427 6271  
## 11 163 445 3779 8988  
## 12 156 525 3410 7696

En Fuente de Notificación pudimos apreciar que la Secretaría de salud siempre ha sido la institución que ha registrado mayor mordeduras de perro en los últimos doce años y que la "SEMAR" es la que menos ha registrado. En estos datos podemos observar una media muy grande al igual que la varianza, probablemente se debe a los valores atípicos que se pudieran registrar. Para eso se harán más análisis para conocer a fondo los datos capturados.

#### 6.1.3 Información De Mes de Ocurrencia

## ANIO MES.MAX MAX MES.MIN MIN MEDIA VAR STD 1st.Q  
## 1 2004 MAR 15317 DIC 7138 11380.250 5310779 2304.513 7138  
## 2 2005 MAY 14234 DIC 7539 11094.333 3544723 1882.743 7539  
## 3 2006 MAY 14972 DIC 7087 11134.000 4174985 2043.278 7087  
## 4 2007 JUL 12079 DIC 6651 9736.000 2382739 1543.612 6651  
## 5 2008 ABR 13389 NOV 7437 9552.417 2852925 1689.060 7437  
## 6 2009 MAR 12100 NOV 6680 8828.417 2649862 1627.840 6680  
## 7 2010 JUN 11305 NOV 6495 8609.000 2308815 1519.479 6495  
## 8 2011 MAR 11655 DIC 6744 9174.500 2654164 1629.160 6744  
## 9 2012 MAY 12643 DIC 5919 9850.917 3540298 1881.568 5919  
## 10 2013 MAY 11267 DIC 6351 8747.000 2183534 1477.679 6351  
## 11 2014 JUL 12271 NOV 8007 9837.750 1768872 1329.990 8007  
## 12 2015 JUN 11939 DIC 6171 9294.083 2829051 1681.978 6171  
## 2nd.Q 3rd.Q 4rd.Q  
## 1 10338.50 11235.5 12247.25  
## 2 10090.25 10990.0 11875.50  
## 3 10498.00 11322.5 11943.75  
## 4 9348.50 9922.5 10247.50  
## 5 8581.75 9172.5 10116.75  
## 6 7861.75 8771.5 9252.75  
## 7 7436.00 8659.0 9328.25  
## 8 8095.75 9058.5 9984.75  
## 9 9387.50 9768.0 10610.00  
## 10 7969.25 8627.0 9399.50  
## 11 8900.50 9508.5 10586.00  
## 12 8358.00 9325.0 10123.75

Para los datos obtenidos en Mes de Ocurrencia podemos apreciar que para el máximo varía con mayor frecuencia el mes siendo: Marzo, Abril, Mayo, Junio, y Julio los meses a través de los años con mayor mordedura de perros. En cambio para el mínimo solo lo comparten Noviembre y Diciembre. También podemos observar como la media va disminuyendo a través de los años, al igual que la desviación estandar, habiendo menor dispersión de los datos con respecto a la media.

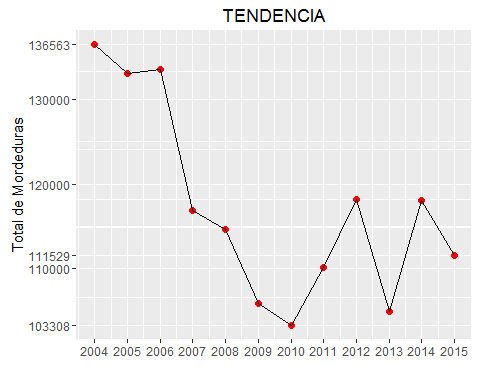
#### 6.1.4 Información De Grupo de Edad

## ANIO EDAD.MAX MAX EDAD.MIN MIN  
## 1 2004 VEINTICINCO.A.CUARENTAYCUATRO 27616 NO.SE.REPORTO.LA.EDAD 82  
## 2 2005 VEINTICINCO.A.CUARENTAYCUATRO 26990 NO.SE.REPORTO.LA.EDAD 111  
## 3 2006 VEINTICINCO.A.CUARENTAYCUATRO 27313 NO.SE.REPORTO.LA.EDAD 77  
## 4 2007 VEINTICINCO.A.CUARENTAYCUATRO 23894 NO.SE.REPORTO.LA.EDAD 76  
## 5 2008 VEINTICINCO.A.CUARENTAYCUATRO 23420 NO.SE.REPORTO.LA.EDAD 75  
## 6 2009 VEINTICINCO.A.CUARENTAYCUATRO 21929 NO.SE.REPORTO.LA.EDAD 174  
## 7 2010 VEINTICINCO.A.CUARENTAYCUATRO 20976 NO.SE.REPORTO.LA.EDAD 51  
## 8 2011 VEINTICINCO.A.CUARENTAYCUATRO 22447 NO.SE.REPORTO.LA.EDAD 55  
## 9 2012 VEINTICINCO.A.CUARENTAYCUATRO 24607 NO.SE.REPORTO.LA.EDAD 103  
## 10 2013 VEINTICINCO.A.CUARENTAYCUATRO 22168 NO.SE.REPORTO.LA.EDAD 169  
## 11 2014 VEINTICINCO.A.CUARENTAYCUATRO 25772 NO.SE.REPORTO.LA.EDAD 245  
## 12 2015 VEINTICINCO.A.CUARENTAYCUATRO 25065 NO.SE.REPORTO.LA.EDAD 72  
## MEDIA VAR STD 1st.Q 2nd.Q 3rd.Q 4rd.Q  
## 1 11380.250 82021526 9056.574 82 5172.50 9149.5 15682.00  
## 2 11094.333 76920144 8770.413 111 5178.25 8907.0 15161.75  
## 3 11134.000 75966876 8715.898 77 5311.50 9210.5 14774.25  
## 4 9736.000 57105106 7556.792 76 4662.75 8239.5 12777.75  
## 5 9552.417 53514025 7315.328 75 4850.50 8152.0 12340.00  
## 6 8828.417 45939554 6777.872 174 4444.25 7568.5 11281.50  
## 7 8609.000 42478573 6517.559 51 4361.25 7390.0 11111.25  
## 8 9174.500 48045037 6931.453 55 4682.75 8106.5 12018.00  
## 9 9850.917 55606726 7456.992 103 5150.75 8789.0 12720.25  
## 10 8747.000 43216589 6573.933 169 4771.00 7890.0 10878.75  
## 11 9837.750 53742041 7330.896 245 5522.50 9323.0 11829.50  
## 12 9294.083 47941516 6923.981 72 5433.75 8953.0 11145.25

Como era de esperarse, al ser el intervalo con mayor número de edades, el intervalo de Veinticinco a Cuarenta y Cuatro es el que registra mayor número de mordeduras. Cuando no se reportó la edad, es comprensible que registrara menores casos, porque implica que en la mayoría de los casos se reporta la edad. podemos notar que la media oscila entre 8,500 y 12,000 pero en la mayoría de los casos vemos media entre 9,000 y 10,000 haciendo el supuesto de que en promedio muerden aproximadamente mordieron a más de 9000 personas en cada grupo de edad

### 6.2 Tendencia de las Mordeduras

Uno de los primeros planteamientos fue: ¿Cómo la incidencia en mordeduras se había comportado a través de los años? y ¿qué tanto había cambiado?

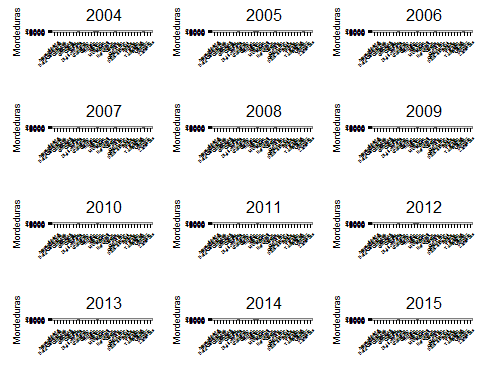


Se creó una gráfica que presentara el total de mordeduras en cada año. Por espacio y limpieza se decidió dejar solo algunos intervalos entre los que destacan el máximo, el mínimo, el primer valor y el último valor.

Con esta gráfica pudimos apreciar que la incidencia de mordeduras de perro del año 2004 a 2015 ha tenido la tendencia a bajar a través de los años en General. Se puden apreciar incrementos a partir del 2010 pero también hubo decrementos marcados como en el 2012 y en 2014. Lo más razonable sería pensar que la incidencia se mantuvo en un nivel bajo a comparación del 2004 que fue el año donde más mordeduras hubo.

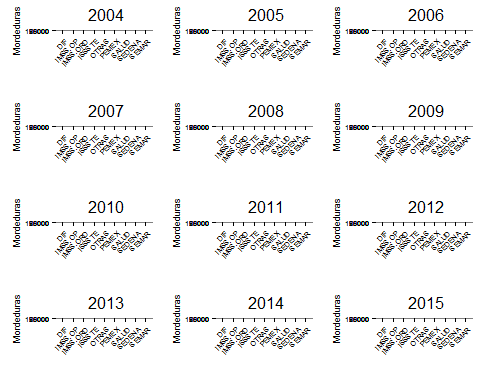
### 6.3 Multigráficas a través de los años

#### 6.3.1 El cambio de las Mordeduras en Cada Estado



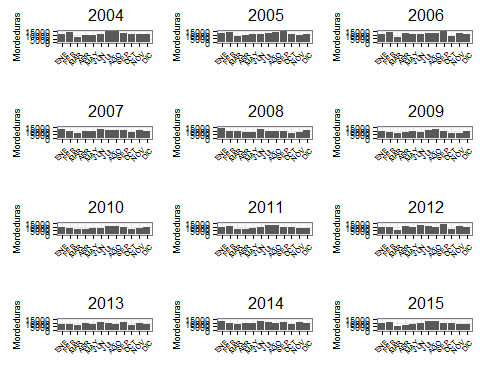
En cada gráfica por año de cada estado y sus mordeduras totales, podemos apreciar como cada año que pasa parace haber un decremento en las mordeduras en todos los estados. También podemos apreciar que, a pesar de que las mordeduras bajaron, los estados tienden a tener la misma relación uno con el otro, esto es, la variabilidad de los estados en cuanto a mordeduras no parece cambiar en gran medida.

#### 6.3.2 El Cambio de las Mordeduras en Cada Fuente de Notificación



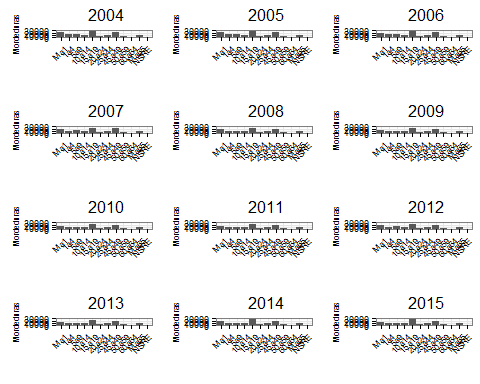
En esta gráfica podemos notar el cambio a través de los años de la incidencia de mordeduras en cada una de la instituciones de salud mencionadas. Podemos apreciar que el cambio en cuanto a las notificaciones de mordeduras de perro no cambian en gran medida. Se puede apreciar que "SALUD" es la que reporta más casos de mordeduras en todos los años. "SALUD" la consideramos un valor atípico ya que por mucho registra mayor cantidad de mordeduras que otras instituciones. También podemos notar como "IMSS.OP" tiende a bajar el registro de mordeduras mientras que "ISSTE" registra cada vez más. Es importante aclarar que en muchas instituciones no se registraron datos es por eso que en la mayoría de las gráficas hay muchas instituciones que no se logran apreciar con gran claridad.

#### 6.3.3 El Cambio de las Mordeduras en Mes de Ocurrencia



En los meses podemos hacer la comparación con la gráfica de tendencia y ver que en efecto, a través de los años va bajando la tendencia. En general se puede apreciar que la tendencia suele ser la misma a través en los meses y la variabilidad de mordedruas de perro en cada mes es muy baja.

#### 6.3.4 El Cambio de las Mordeduras en Grupo de Edad

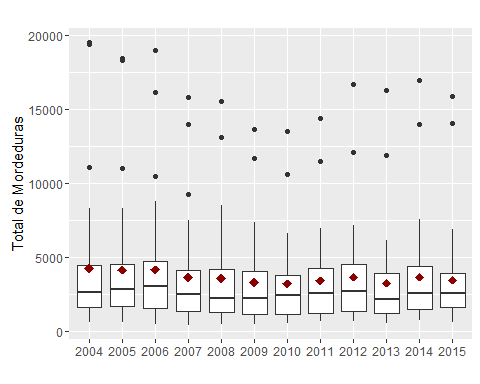


A medida que van cambiando los años suele haber un decremento en las mordeduras de perro en cada grupo de edad. De antemano, notamos que el grupo de edad que más mordeduras tuvo fue el de "25-44" ya que es el grupo con un mayor intervalo de edad por lo tanto pueden haber más personas dentro de este intervalo. Uno de los puntos que consideramos importantes en este análisis es la cantidad de perros que muerden a niños muy chicos, ya que se encuentra en el tercer lugar en cuanto a mordeduras. Creemos que se podrían tomar medidas para responsabilizar a los padres dado que las mordedruas de perro ocurren con gran frecuencia a edades muy chicas.

### 6.4 Boxplots

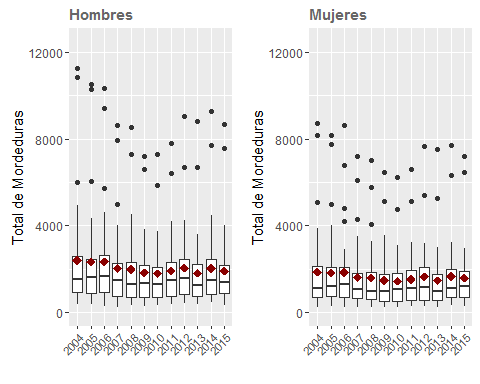
Para cada Boxplot la media esta representada por un rombo rojo.

#### 6.4.1 Boxplot del Total de Mordeduras a Través de los Años de Cada Estado



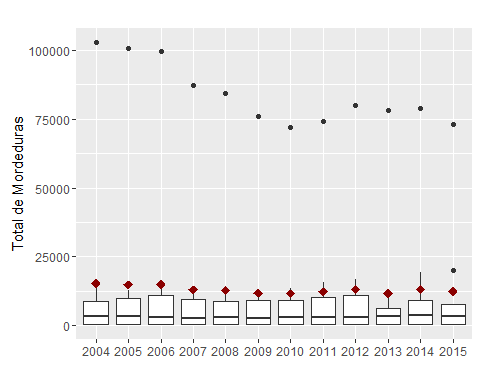
En este boxplot se representa el total de mordeduras que hubo a través de los años para cada estado. Por medio del Boxplot se pudo deducir que en todos los años la media estaba por arriba de la mediana. Esto es evidencia de que la distribución de la incidencia en mordeduras se encuentra sesgada a la derecha en todos los años. Con respecto a valores atípicos, destaca el "Distrito Federal" y el Estado de México, en este caso "México". Ambos registran el mayor número de mordedruras de todos los estados. La media generalmente se encuentra alrededor de 5000 y 2500, aproximadamente muerden esta cantidad de perros a personas en los estados. Podríamos formular hipótesis al respecto, probablemente exista una relación directa entre el número de personas que habitan una región y el número de perros, recordemos que estos son animáles domésticos. Una de nuestras limitantes es no conocer el total de personas porque se podría hacer un estimado de cuantas personas del total de la población son mordidas por un perro.

##### 6.4.1.1 Boxplot del Total de Mordeduras a Través de los Años de Cada Estado para Hombres y Mujeres



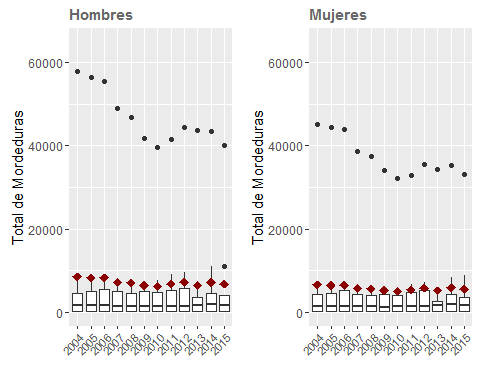
En este boxplot se hace la separación de hombres y mujeres.Podemos apreciar, al igual que en la observación general, la media esta por encima de todas las medianas generando sesgos en ambos boxplots. Un punto a resaltar, es que en cuanto a Hombres y Mujeres se puede llegar a la conclusión que en general muerden más a hombres que a mujeres en todos los Estados. A pesar de que sus respectivos valores atípicos se encuentran en mayor diferencia, comparando el del hombre con el de la mujer, las medias y las medianas se encuentran por el mismo intervalo y no es gran la dispersión de estos, con esto podemos conlcuir que podría no afectar ser hombre o mujer pueden tener la misma probabilidad de ser mordidos.

#### 6.4.2 Boxplot del Total de Mordeduras a Través de los Años de Fuente de Notificación



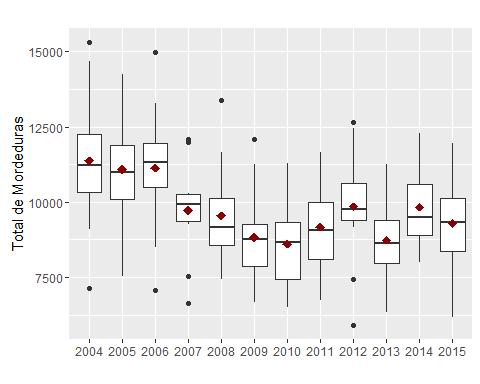
El boxplot de Fuente de Notificación permite visualizar que la media supera el tercer cuartil, y se hace demasiado sesgada a la derecha. Esto ocurre porque en Fuente de Notificación existen varios "Sin registros" por parte de las instituciones y como la institución "SALUD" contiene la mayoría de los casos desplaza la media por arriba de la mediana, convirtiéndose "SALUD" en un valor extremadamente atípico.

##### 6.4.2.1 Boxplot del Total de Mordeduras a Través de los Años de Fuente de Notificación para Hombres y Mujeres

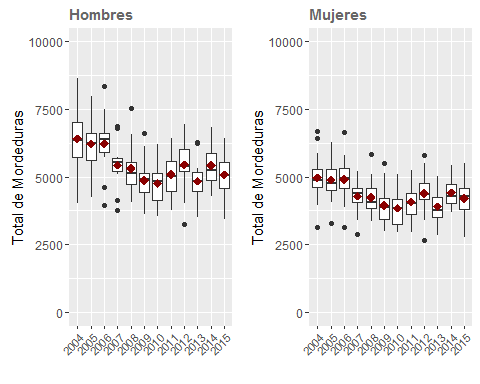


Consideramos los boxplots de hombres y mujeres para fuente de notificación para entender cómo se estaban registrando las mordeduras para ambos géneros. Pudimos apreciar que las instituciones registran más hombres que mujeres con mordeduras de perro. Aun así, al igual que por Estados, la mediana y la media no tienen un gran cambio en hombre mujer.

#### 6.4.3 Boxplot del Total de Mordeduras a Través de los Años de Fuente de Mes de Ocurrencia

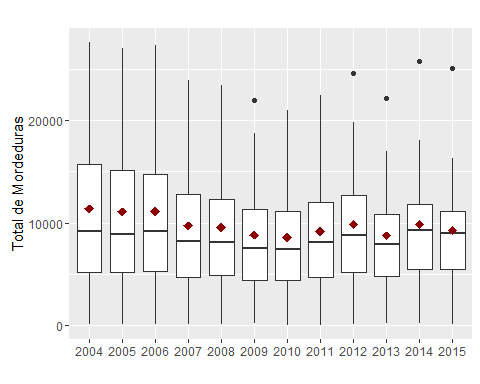
 En Mes de Ocurrencia se presentan datos más dispersados que en los pasados boxplots. Hay sesgos a las derecha como a la izquierda, e incluso se acercan varios años a la simetría, como en 2015. En algunos años no existen valores atípicos. Si se compara con la gráfica de Tendencia se concluye que es razonable que la media vaya bajando a través de los años ya que previamente se encontró evidencia que la cantidad mordeduras disminuyó con en el transcurso del tiempo.

##### 6.4.3.1 Boxplot del Total de Mordeduras a Través de los Años de Mes de Ocurrencia para Hombres y Mujeres



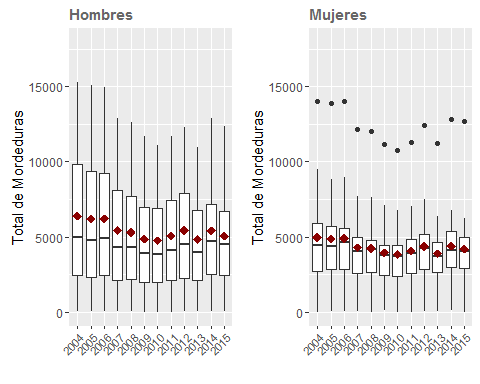
Al observar el boxplot de hombres y mujeres se encuentra evidencia de que los perros muerden más a los hombres que a las mujeres. En este caso notamos una mayor desviación de la media y la mediana de hombres con respecto a mujeres, implicando que muerden en mayor medida a los hombres cada año, ya que si se esta haciendo por mes el promedio va ligado estrictamente al año mencionado.

#### 6.4.4 Boxplot del Total de Mordeduras a Través de los Años de Grupo de Edad



En el boxplot de Grupo de Edad se aprecia un sesgo a la derecha en todos los años, pero se puede ver también que el sesgo ha estado disminuyendo através de los años, esto hace la media más cercana a la mediana. Con lo anterior se puede argumentar que ha habido mayor similitud en las mordeduras de grupos de edad a través de los años. Uno de los problemas que siempre registrará el gurpo de edad es que existe un intervalo demasiado grande que abarca a más persona que en otro.

##### 6.4.4.1 Boxplot del Total de Mordeduras a Través de los Años de Grupo de Edad para Hombres y Mujeres



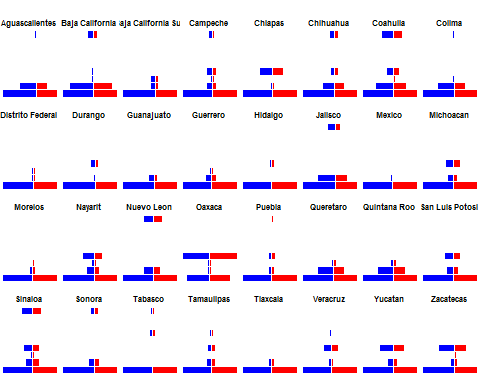
En el boxplot de Hombres notamos un sesgo a la derecha mucho mayor que en el de las mujeres. En el de las mujeres apreciamos un mayor acercamiento a la simetría que en el caso de los hombres.

### 6.5 Pirámides

Cada pirámide representa el total de mordeduras por estado dependiendo de la fuente utilizada: Fuente de Notificación, Mes de Ocurrencia o Grupo de Edad.

En azul esta representado el total de mordeduras de hombres. En rojo esta representado el total de mordeduras de mujeres.

#### 6.5.1 Pirámide de Fuente de Notificación

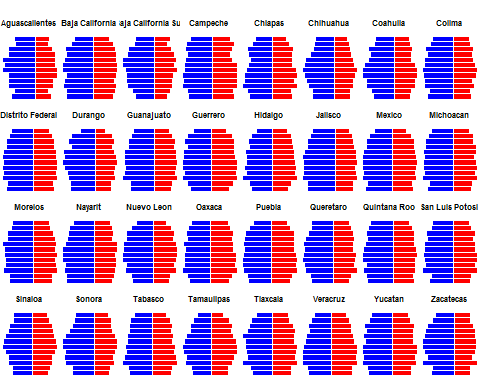


Estas gráficas permiten tener una idea visual de qué instituciones concentran la mayor cantidad de reportes de mordeduras de perro en cada uno de los estados. Se omitieron las etiquetas que indican la institución que representa cada escalón por cuestiones de falta de espacio, de haber incluido dichas etiquetas hubiese resultado en un exceso de información para el lector y por lo tanto mayor dificultad para interpretarlas. Hacemos mención en orden ascendente: "SALUD", "IMSS.ORD","ISSSTE","IMSS.OP","DIF","PEMEX", "SEDENA","SEMAR","OTRAS".

Llama la atención ver repetidamente varios escalones vacios, esto se debe a dos razones. La primera distintas instituciones, como el caso de PEMEX y SEMAR, no se encuentran en todos los estados; por lo tanto el número de casos que reportan, por ejemplo en los estados del centro del país, tienden a ser cero. La segunda razón es que se está haciendo una comparación en proporciones, en el código se fijaron dichos límites de proporciones como el máximo de casos reportados tanto en masculino como en femenino, para cada uno de los estados. Pese a lo anterior, la información es consistente al revelar nuevemente que "SALUD" es la institución que predomina en todos los estados en reportar casos de mordeduras. El único caso en que dicha institución no predomina es en Tamaulipas, donde "IMMS.OP" es la principal.

#### 6.5.2 Pirámide de Mes de Ocurrencia

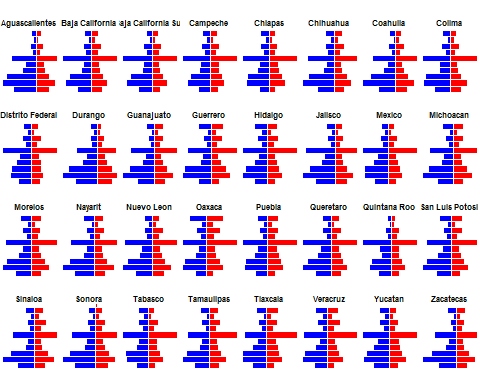
Definimos cada uno de los escalones en orden ascendente: "ENERO","FEBRERO", "MARZO","ABRIL","MAYO","JUNIO","JULIO","AGOSTO","SEPTIEMBRE","OCTUBRE","NOVIEMBRE","DICIEMBRE".



Notamos que la distribución de merdeduras de perro por mes de ocurrencia dividido en hombres y mujeres, tiene un comportamiento similar a una distribución uniforme en la mayoría de los casos. Se reafirma la hipótesis de que los perros muerden por igual en cualquier mes del año y que a pesar de que los hombres son más propensos a sufrir mordedsuras de perros, no existe una gran diferencia entre el total de modeduras en mujer y en hombre.

#### 6.5.3 Pirámide de Grupo de Edad

Definimos los escalones en orden ascendente: ">1", "1-4", "5-9", "10-14", "15-19", "20-24", "25-44", "45-49", "50-59", "60-64", ">65", "Ign".



Las pirámides de cada uno de los estados tienen formas similares, al igual que en los boxplots notamos sesgos a la derecha (arriba). Reafirmamos la hipótesis antes mencionada de que la incidencia en mordeduras de perro se concentra en el grupo de edad de 25 a 44 años, ya que es un grupo demasiado grande en lo que respecta a la probabilidad de que un evento caiga en ese rango. Creemos que el sector salud debe prestar especial atención en niños de edades entre uno y nueve años, ya que son una población muy vulnerable, sobre todo por la amenaza de adquirir alguna enfermedad transmitida por las mordeduras.

## 7. Conclusiones

Después de observar, analizar y graficar los datos obtenidos, se pudo conlcuir: Primero, los datos de cualquiera de las fuentes (Fuente de Notificación, Mes de Ocurrencia o Grupo de Edad) presentan una gran similitud a través de los años, podrán variar y disminuir las mordeduras perro en cada una de las variables orginales, siempre oscilan en el mismo intervalo. Esto nos obliga a dudar de la veracidad de los datos, nos hace preguntarnos si en verdad se estan capturando los datos de manera eficiente y real. En caso de que estemos equivocados tenemos que conlcuir que las Mordeduras de Perro presentan un patrón muy común.Se pueden utilizar métodos numéricos para crear aproximaciones a los patrones y hacer buenas estimaciones de los próximos años.

Otra de nuestras conclusiones nace de la captura de los datos. En fuente de Notifiación, prácticamente solo hay una institución que esta registrando las mordeduras de perro, también creemos que "OTRAS" puede llegar a resultar muy ambiguo y con motivos a dudar si en "OTRAS" se incuyen a todas las instituciones públicas como tanto privadas. En Grupo de Edad, se deben cambiar los intervalos de edad porque al momento de hacer cualquier análisis estadístico se alteraran las conlcusiones si se tiene un intervalo de 19 años.

Creemos que México debe ser educado en cuanto a la captura de información, México no se esta actualizando en las capturas de información y esto esta provocando análisis dudosos y poco nutritivos. Si los análisis se estan haciendo con datos falsos o con datos faltantes se estan haciendo análisis falsos y análisis incompletos. Cada país toma decisiones con base a análsis estadísticos, porque estos se encargan de presentar pronósticos y aproximaciones de la realidad, y con esto dichas decisiones más acertadas a lo demandado por la realidad. Si México esta capturando erróneamente sus datos esta produciendo datos falsos y esto implicaría malas decisiones, y un mal empleo de los recursos.

Creemos que es tiempo de revisar y cuestionar los métodos utilizados para capturar información, ya que los estadísticos tiene como obligación presentar datos verídicos como los menciona Stephen Few y como lo establece "The Actuarial Standard of Practice" Se debe ser crítico con los datos y mucho más con nuestro trabjo como lo menciona Edward Tufte, por eso en este trabajo se trata de ser muy crítico y reflejar la realidad de los datos capturados.