

# Diabetes Melitus Tipo II

*Sebastian Morales Peralta 000124317*

*27 de marzo de 2015*

## REPORTE TECNICO

### I. Introduccion

Este reporte busca presentar, de forma general, el fenomeno de la Diabetes melitus tipo II (no insulino-dependiente) en Mexico. Los datos contenidos abarcan el periodo 2003 a 2013 y son publicados por la Secretaria de Salud en la pagina:

<http://www.epidemiologia.salud.gob.mx/anuario/html/anuarios.html>

En esta seccion se encuentra una breve descripcion de los datos para pasar, mas adelante, a la definicion del problema.

La base de datos esta compuesta por 3 tablas principales, presentadas a continuacion, describen los casos de diabetes reportados de 2003 a 2013. La informacion esta distribuida en dos generos sexuales (masculino y femenino) en las 32 entidades del pais.

FUENTE.

Informacion dividida por las distintas instituciones que reportaron los casos de diabetes.

```
##          estado  genero fecha fuente casos
## 1      Agascalientes General  2003  salud  1859
## 2      Baja California General  2003  salud  2443
## 3 Baja California Sur General  2003  salud   646
## 4          Campeche General  2003  salud  1188
## 5          Coahuila General  2003  salud  1346
## 6          Colima General  2003  salud   802
```

```
## [1] "salud"      "imss.ord" "issste"   "imss.op"  "dif"      "pemex"
## [7] "sedena"     "semar"    "otras"
```

MES.

La informacion se encuentra dividida por cada mes del anio.

```
##          estado  genero fecha mes casos
## 1      Agascalientes General  2003 ene   268
## 2      Baja California General  2003 ene  1049
## 3 Baja California Sur General  2003 ene   171
## 4          Campeche General  2003 ene   206
## 5          Coahuila General  2003 ene  1299
## 6          Colima General  2003 ene   174
```

EDAD.

Presenta los casos de diabetes reportados en los distintos grupos de edad, donde ign se refiere a aquellos individuos que no fueron clasificados en un grupo de edad.

```
##          estado  genero fecha edad casos
## 1      Agascalientes General  2003  <1    0
## 2      Baja California General  2003  <1    0
## 3 Baja California Sur General  2003  <1    0
## 4          Campeche General  2003  <1    0
## 5          Coahuila General  2003  <1    2
## 6          Colima General  2003  <1    0

## [1] "<1"    "1a4"    "5a9"    "10a14"  "15a19"  "20a24"  "25a44"  "45a49"
## [9] "50a59"  "60a64"  "65+"    "ign"
```

Nota: se omitieron la mayoría de los acentos en el presente reporte, debido a que eran cambiados por “??” automáticamente por el sistema.

## II. Definición del Problema.

El principal objetivo de este reporte es presentar el fenómeno de la diabetes melitus tipo II en Mexico, por lo cual iniciaremos haciendo las preguntas mas generales, hasta llegar a responder las preguntas mas especificas que nuestros datos nos permitan.

1. Es importante conocer como se distribuyen todos nuestros casos reportados de diabetes entre hombres y mujeres de la poblacion.
2. Cual es el estado de la republica que presenta mayores casos de diabetes en Mexico; asi como aquel estado que presenta los menores casos.
3. Presentar la tendencia, a lo largo del tiempo, de los casos reportados en el periodo 2003 - 2013, ??es creciente o decreciente?
4. ??En que año se reporto el mayor y menor numero de casos de diabetes en Mexico?
5. ??Cual es el comportamiento, de los casos reportados, en las distintas edades de la poblacion? ??En que grupo de edad se encuentra el mayor numero de casos de diabetes?
6. ??Cuales son las instituciones en la republica donde se han reportado el mayor y menor numero de casos?
6. ??Como es el comportamiento mensual de los casos reportados? ??Existe algun tipo de estacionalidad?
7. Cual es el comportamiento de los casos en hombre y mujeres a lo largo del tiempo y en cada uno de los estados de la republica?
8. Cual es el comportamiento de los distintos grupos de edad a lo largo del tiempo y en cada una de las entidades del país?

Cada una de estas preguntas ser respondida mediante un analisis exploratorio, el cual nos permita entender el comportamiento de los casos de diabetes en la poblacion mexicana a lo largo del tiempo, por lo que se incluyan datos de promedios y varianzas para comenzar a “visualizar” que tan solido sera el analisis en la etapa inferencial (no presentada en el actual reporte).

## III. Obtencion y Descripción de los Datos.

Una base de datos “ideal” que podria responder todas las preguntas de manera puntual seria aquella que contuviera los siguientes campos o variables:

```
## [1] "estado" "genero" "fecha" "mes"    "fuente" "edad"  "casos"
```

En la cual, las variables serian de los siguientes tipos y escalas:  
ESTADO.

Esta variable es de tipo categorica y da cuenta de cada uno de los estados de la republica mexicana.

```
## [1] "Aguscalientes"      "Baja California"      "Baja California Sur"
## [4] "Campeche"           "Coahuila"             "Colima"
## [7] "Chiapas"            "Chihuahua"            "Distrito Federal"
## [10] "Durango"            "Guanajuato"           "Guerrero"
## [13] "Hidalgo"            "Jalisco"              "Mexico"
## [16] "Michoacan"          "Morelos"              "Nayarit"
## [19] "Nuevo Leon"         "Oaxaca"               "Puebla"
## [22] "Queretaro"          "Quintana Roo"         "San Luis Potosi"
## [25] "Sinaloa"            "Sonora"               "Tabasco"
## [28] "Tamaulipas"         "Tlaxcala"             "Veracruz"
## [31] "Yucatan"            "Zacatecas"
```

GENERO.

El sexo de de los individuos reportados con diabetes melitus tipo II. Es una variable categorica.

```
## [1] "General" "Mujeres" "Hombres"
```

FECHA.

Se refiere al anio en el que se presentaron los casos de diabetes, es una variable numerica, de tipo entero, que se encuantra entre los rangos 2003 y 2013.

```
## [1] 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
```

MES.

El mes del anio en el que se reportaron los casos, es una variable de tipo caracter que contiene cada uno de los meses del anio.

```
## [1] "ene" "feb" "mar" "abr" "may" "jun" "jul" "ago" "sep" "oct" "nov"
## [12] "dic"
```

FUENTE.

Es una variable categorica que permite saber el numero de casos reportados por algunas instituciones del pais.

```
## [1] "salud"      "imss.ord" "issste"    "imss.op"  "dif"      "pemex"
## [7] "sedena"     "semar"     "otras"
```

EDAD.

Variable categorica que agrupa distintos rangos de edades que presentan casos de diabetes. En este caso, la categoria "ign" se refiere a todas aquellas personas que no fueron asignadas a algun grupo de edad.

```
## [1] "<1"      "1a4"     "5a9"     "10a14"   "15a19"   "20a24"   "25a44"   "45a49"
## [9] "50a59"   "60a64"   "65+"     "ign"
```

CASOS.

Es la variable dependiente, es de tipo numerica y se refiere al numero de personas que reflejan un caso de diabetes melitus tipo II (no insulino dependiente). Para este reporte, la variable se encuentra en los siguientes rangos:

```
##          min    max
## casos    0 28024
```

Este reporte contiene todas estas variables, sin embargo, no se tienen integradas en una sola base de datos, por lo que no es posible dar respuestas tan específicas como por ejemplo:

Cuántos casos de hombres con diabetes se reportaron en abril de 2008 en el issste en Durango?

Ya que la información actual puede dar cuenta de los hombres reportados en abril de 2008 y los hombres reportados en el 2008 por parte del issste pero a lo largo de todo el año.

ej.

```
##          estado  genero fecha mes casos
## 3690 Durango Hombres 2008 abr   269
```

```
##          estado  genero fecha fuente casos
## 2986 Durango Hombres 2008 issste   566
```

Por lo tanto, utilizaremos las tres bases por separado (fuente, mes, edad) para responder de las preguntas más generales hasta la más particular que sea posible.

#### IV Calidad de los Datos.

En esta sección se mostrará la calidad de los datos presentados, es decir, que los datos de cada una de nuestras tablas (fuente, mes, edad) sean los mismos para cada año, género, y estado. Esto con el fin de corroborar que los datos que estamos utilizando coinciden, independientemente de la tabla que estemos utilizando, lo cual nos permite tener solidez en el análisis comparativo.

Primero mostraremos que, para cada tabla, los datos son los mismos para cada año y cada género. Por fuente.

##	Fecha	General	Hombres	Mujeres	Total	Diferencia
## 1	2003	380332	147171	233161	380332	0
## 2	2004	404562	156916	247646	404562	0
## 3	2005	397387	152621	244766	397387	0
## 4	2006	394360	152836	241524	394360	0
## 5	2007	404770	160068	244702	404770	0
## 6	2008	396374	158992	237382	396374	0
## 7	2009	426802	177825	248977	426802	0
## 8	2010	420032	173507	246525	420032	0
## 9	2011	427125	176891	250234	427125	0
## 10	2012	419493	173725	245768	419493	0
## 11	2013	397192	166220	230972	397192	0

Por mes.

##	Fecha	General	Hombres	Mujeres	Total	Diferencia
## 1	2003	380332	147171	233161	380332	0
## 2	2004	404562	156916	247646	404562	0
## 3	2005	397387	152621	244766	397387	0
## 4	2006	394360	152836	241524	394360	0
## 5	2007	404770	160068	244702	404770	0
## 6	2008	396374	158992	237382	396374	0

## 7	2009	426802	177825	248977	426802	0
## 8	2010	420032	173507	246525	420032	0
## 9	2011	427125	176891	250234	427125	0
## 10	2012	419493	173725	245768	419493	0
## 11	2013	397192	166220	230972	397192	0

Por edad.

##	Fecha	General	Hombres	Mujeres	Total	Diferencia
## 1	2003	380332	147171	233161	380332	0
## 2	2004	404562	156916	247646	404562	0
## 3	2005	397387	152621	244766	397387	0
## 4	2006	394360	152836	241524	394360	0
## 5	2007	404770	160068	244702	404770	0
## 6	2008	396374	158992	237382	396374	0
## 7	2009	426802	177825	248977	426802	0
## 8	2010	420032	173507	246525	420032	0
## 9	2011	427125	176891	250234	427125	0
## 10	2012	419493	173725	245768	419493	0
## 11	2013	397192	166220	230972	397192	0

Nota: Total = Hombres + Mujeres

Nota: Diferencia = Total - General

En total, para cada genero y en cada año, podemos observar que los datos son los mismos para cada una de las tablas (fuente, mes y edad). Y en un primer analisis nos estaria diciendo el total de hombres y mujeres reportados con un caso de diabetes melitus tipo II, asi como el total de personas reportadas en el periodo.

```
## Hombres Mujeres    Total
## 1796772 2671657 4468429
```

```
## Hombres Mujeres    Total
## 1796772 2671657 4468429
```

```
## Hombres Mujeres    Total
## 1796772 2671657 4468429
```

## V. Transformacion de los Datos.

Los datos que tenemos estan listos para comenzar el analisis exploratorio, simplemente extraeremos las variables de lo general a lo particular, dependiendo de la o las preguntas que se deban contestar.

ej. Casos de hombres reportados por el DIF.

```
hombresDif <- subset(fuente, genero == "Hombres" & fuente == "dif")$casos
sum(hombresDif)
```

```
## [1] 27081
```

Sin embargo, en esta seccion se explica el proceso que se llevo a cabo para obtener los datos listos para utilizarse.

## V.I. Extraccion de los Datos.

Primero, se llevo a cabo la extraccion de los datos desde la pagina de la Secretaria de Salud:  
<http://www.epidemiologia.salud.gob.mx/anuario/html/anuarios.html>  
Fueron descargados 99 archivos PDF que contienen la informacion de los casos de diabetes reportados para hombres, mujeres y a nivel general para los años 2003 a 2013, dependiendo de la categoria seleccionada: Grupo de Edad, Tipo de Fuente y Mes.

## V.II. Cambio de Formato (PDF a CSV).

Para poder hacer uso de los datos extraidos, el formato de los archivos debe ser de texto plano, es decir, .txt o .csv. Para hacer el cambio de formato de PDF a CSV utilizo el programa TABULA (<http://www.tabula.technology>).

De los archivos PDF fueron transformados 99 archivos CSV, a cada archivo le fue asignado un nombre especificando el genero y el año de los datos contenidos y despues guardados en carpetas dependiendo de la categoria de los datos(Tipo de Fuente, Grupo de Edad o Mes).

ej.

“PROYECTO/DATOS/BRUTOS/CSV/MES/G-2009” Se refiere al reporte general de 2009 por mes.

“PROYECTO/DATOS/BRUTOS/CSV/FUENTE/M-2003” se refiere al reporte de mujeres en 2003 por tipo de fuente.

Esto con el fin de homologar los nombres y generar una secuencia para leer, procesar y compilar los datos mediante codigo R.

## V.III. Limpieza y Compilacion de Datos.

Mediante codigo R se realiza la lectura, limpieza y compilacion de los datos para, finalmente, generar tres nuevos archivos con la informacion lista para realizar los respectivos analisis.

La lectura consiste en cargar los 33 archivos csv de cada una de las categorias para hombres, mujeres y general durante el periodo 2003 a 2013.

La limpieza de los datos consiste en la homologacion de titulos, orden de las columnas, simplificar los nombres de las variables categoricas, como por ejemplo:

“San Luis Potos??” en “San Luis Potosi”.

Tambien se quitan los espacios entre el numero de casos para que la variable pueda ser tomada en cuenta como numerica, ejemplo:

“2 45” a 245.

La compilacion de los datos es simplemente integrar los 33 archivos de cada categoria en una sola tabla.

Para mayor informacion al respecto se puede revisar el codigo en los siguientes archivos:

PROYECTO/CODIGO/PRELIMINAR/Limpieza.R

PROYECTO/CODIGO/PRELIMINAR/Limpieza.RData

## VI. Analisis Exploratorio.

### Analisis General.

En genera, se ha registrado una mayor cantidad de diabetes melitus tipo II en mujeres (periodo 2003-2013)

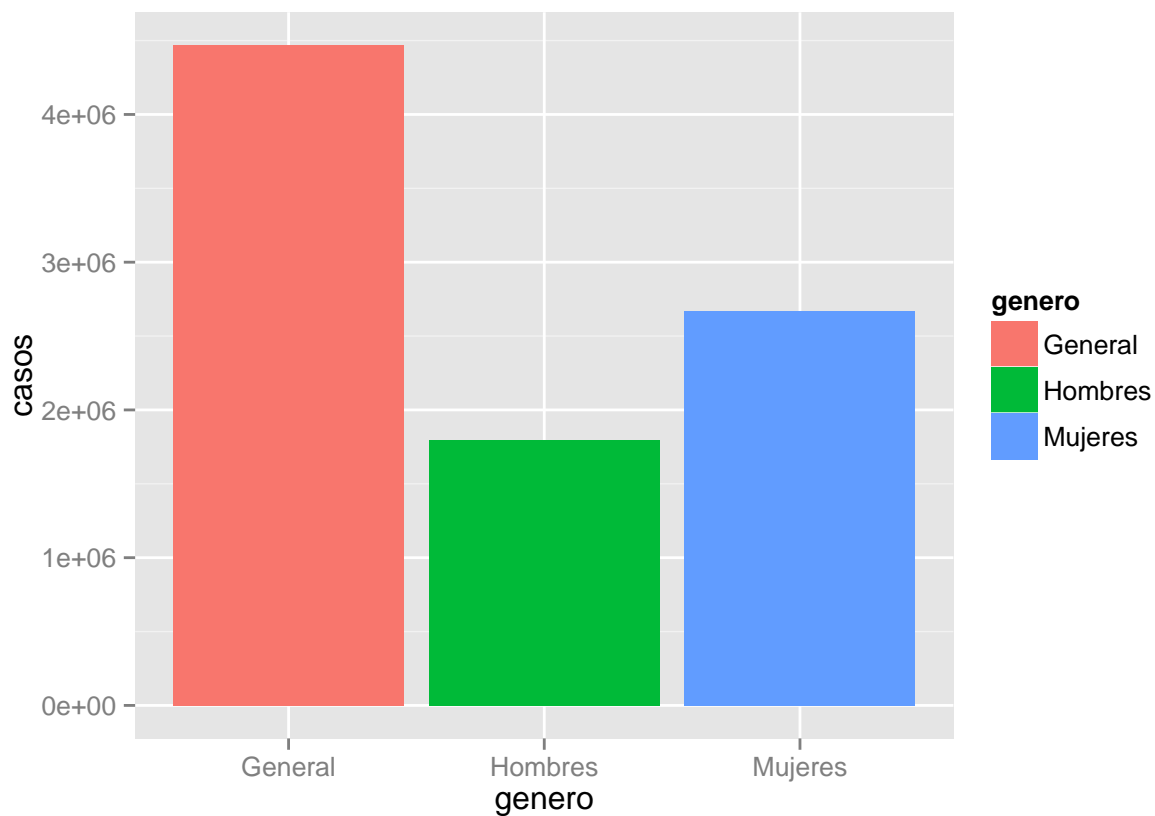
```
##      Hombres Mujeres  Total
## Casos 1796772 2671657 4468429
```

El porcentaje en hombres y en mujeres es el siguiente:

```
##      Hombres  Mujeres
## % 40.21037 59.78963
```

```
## pdf
## 2
```

```
## pdf
## 2
```



En el año 2003 se registro el menor número de casos de diabetes en la población general, mientras que 2011 fue el año con más casos.

En hombres, el menor número se registro en 2003 y el mayor en 2009.

En mujeres, el menor número se registro en 2013 y el mayor en 2011.

```
##      Hombres Mujeres General
## min    2003    2013    2003
## max    2009    2011    2011
```

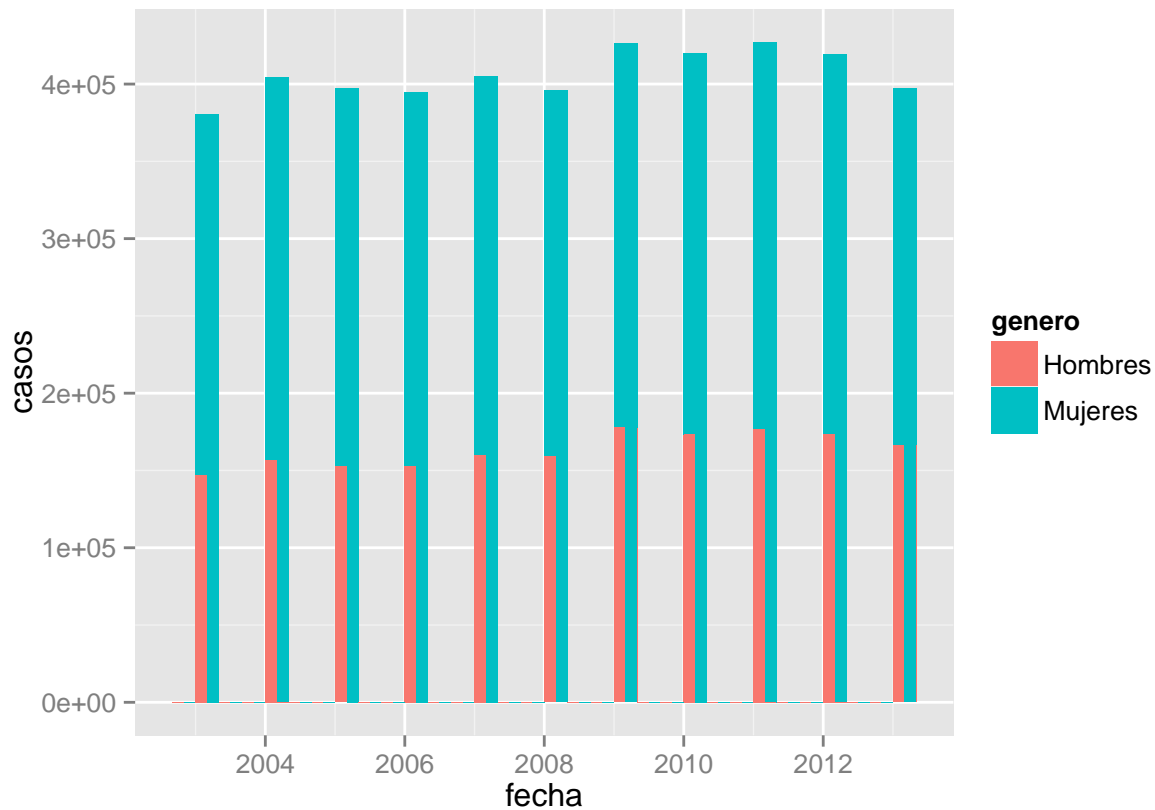
El promedio de casos registrados por año en hombres, mujeres y a nivel general es el siguiente:

```
##      General  Mujeres  Hombres
## 1 406220.8 242877.9 163342.9
```

La tendencia a lo largo del tiempo en hombres y mujeres:

```
## pdf
## 2
```

```
## pdf
## 2
```



Podemos concluir, de este analisis, que la tendencia a nivel general es creciente.

Los estados que reportan el mayor y menor numero de casos son los siguientes:

```
##      Hombres Mujeres General
## min  Colima  Colima  Colima
## max  Mexico  Mexico  Mexico
```

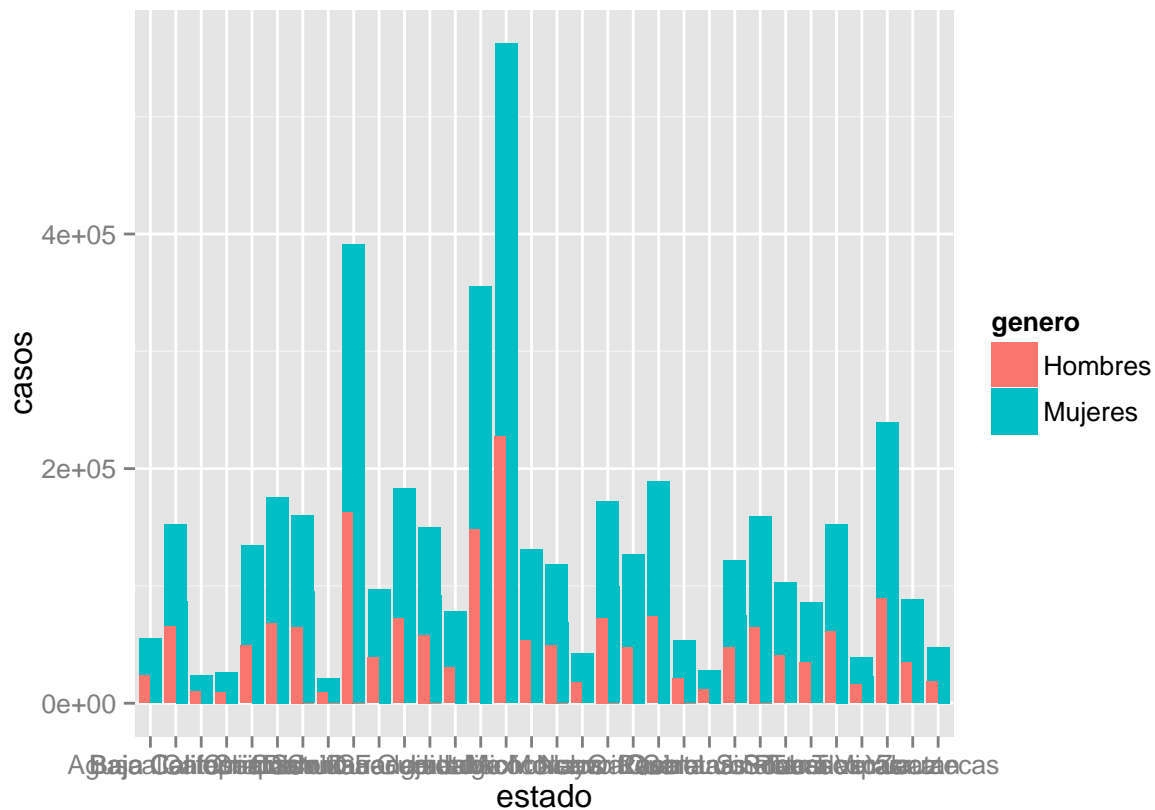
El porcentaje de casos reportados por estado de la republica en el periodo 2003-2013:



##	Estado	General
## 15	Mexico	12.5940235
## 9	Distrito Federal	8.7529644
## 14	Jalisco	7.9608739
## 30	Veracruz	5.3658903
## 21	Puebla	4.2377757
## 11	Guanajuato	4.0979951
## 8	Chihuahua	3.9198340
## 19	Nuevo Leon	3.8524502
## 5	Coahuila	3.5855331
## 25	Sinaloa	3.5653470
## 2	Baja California	3.4163237
## 28	Tamaulipas	3.4089162
## 12	Guerrero	3.3625241
## 7	Chiapas	3.0118863
## 16	Michoacan	2.9432492
## 20	Oaxaca	2.8333000
## 24	San Luis Potosi	2.7389044
## 17	Morelos	2.6486266
## 26	Sonora	2.3040984
## 10	Durango	2.1687488
## 31	Yucatan	1.9900730
## 27	Tabasco	1.9218387
## 13	Hidalgo	1.7534798
## 1	Aguscalientes	1.2357811
## 22	Queretaro	1.1997953
## 32	Zacatecas	1.0701524
## 18	Nayarit	0.9651490
## 29	Tlaxcala	0.8764825
## 23	Quintana Roo	0.6330189
## 4	Campeche	0.5819271
## 3	Baja California Sur	0.5290898
## 6	Colima	0.4739473

## pdf  
## 2

## pdf  
## 2



En ningún estado es mayor el número de casos en hombres que en mujeres.

### Analisis por Tipo de Fuente

El mayor número de casos reportados por las instituciones del país se dio de manera siguiente:

```
## estado genero fecha fuente casos
## 303 Mexico General 2012 salud 28024

## estado genero fecha fuente casos
## 1039 Mexico Hombres 2013 salud 10869

## estado genero fecha fuente casos
## 655 Mexico Mujeres 2012 salud 17423
```

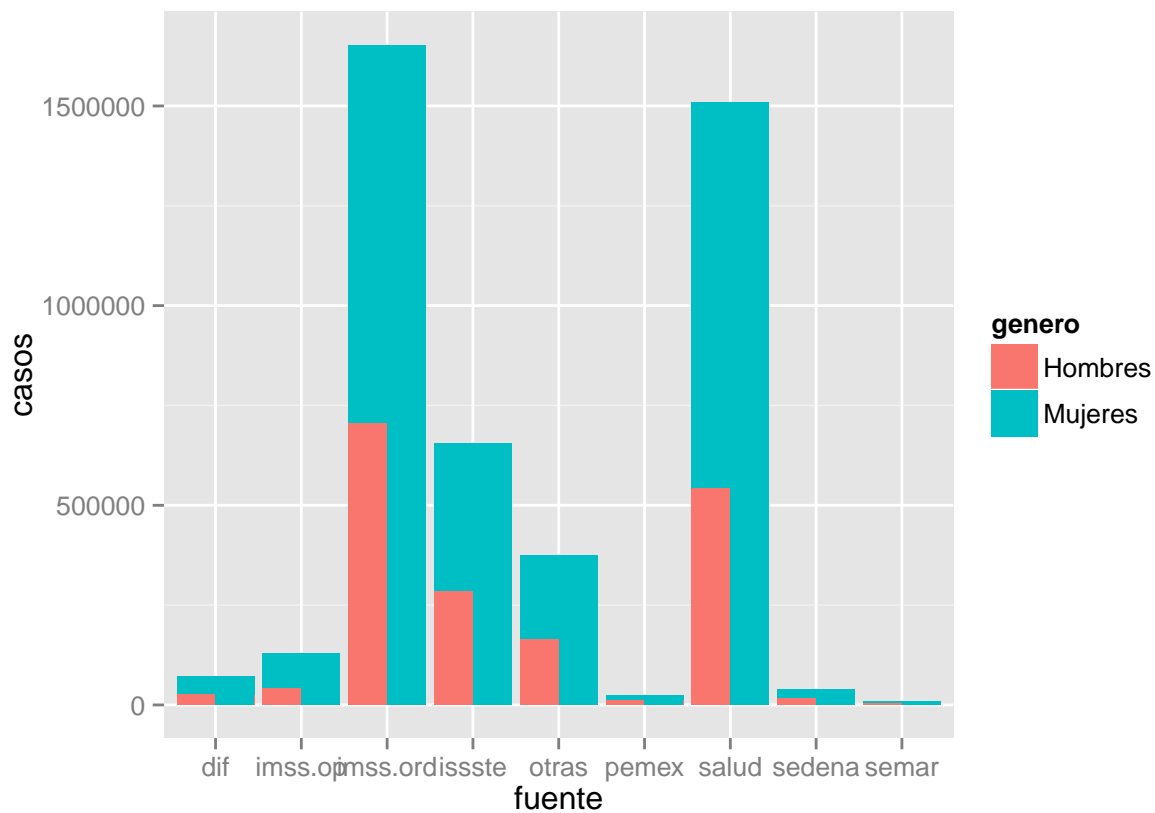
En general, el mayor número de casos de diabetes han sido reportados por la Secretaría de Salud. El porcentaje de reportes de diabetes se distribuye por cada una de las instituciones de la siguiente manera:

```
## Fuente General
## 2 imss.ord 36.9978800
## 1 salud 33.7964864
## 3 issste 14.6857430
## 9 otras 8.3896600
## 4 imss.op 2.8813035
## 5 dif 1.6263882
## 7 sedena 0.8615780
```

```
## 6 pemex 0.5385562
## 8 semar 0.2224048
```

```
## pdf
## 2
```

```
## pdf
## 2
```



A nivel general, los casos tienen el siguiente comportamiento:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         0      22      290    1602    1789    28020
```

Varianza.

```
## [1] 8249738
```

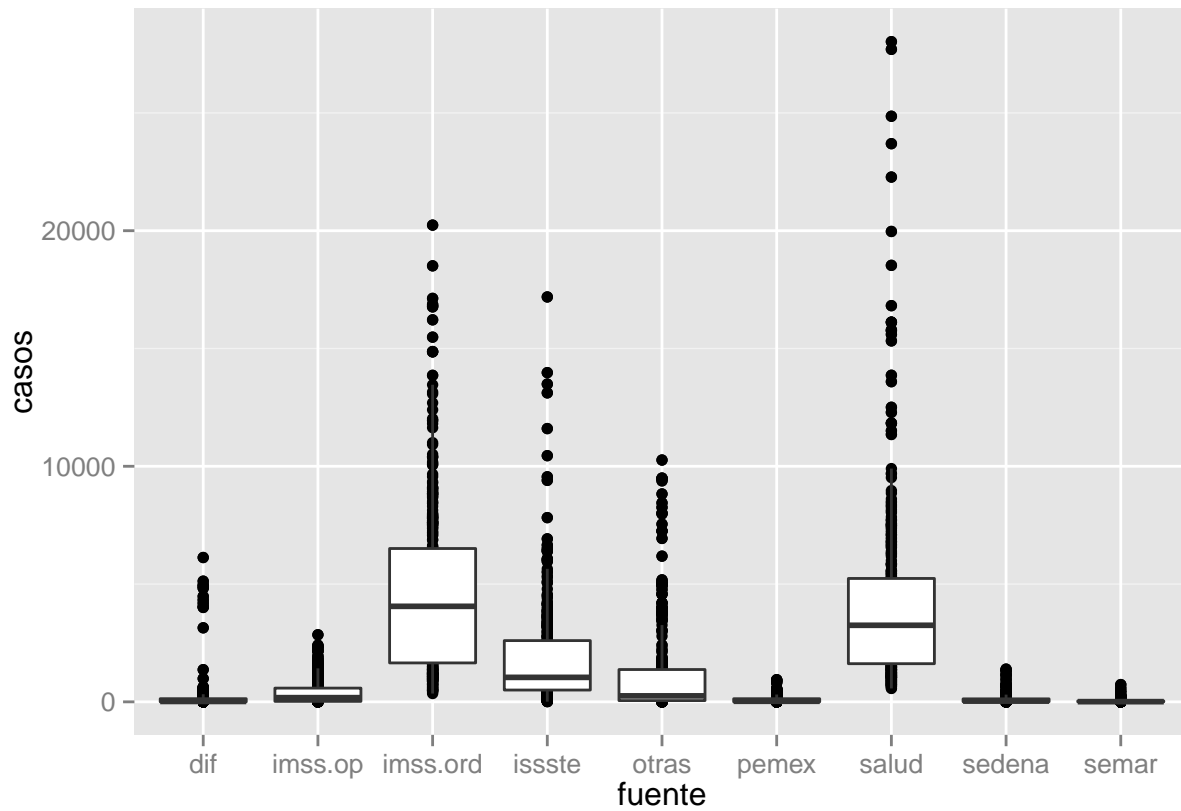
Desviacion estandar.

```
## [1] 2872.236
```

Entonces podemos decir que, a nivel general los casos se encuentran cerca de la media. Sin embargo existen muchos casos atípicos.

```
## pdf
## 2
```

```
## pdf
## 2
```



En hombres, el comportamiento es el siguiente:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0      8.0   119.0   644.2  721.0 10870.0
```

Varianza.

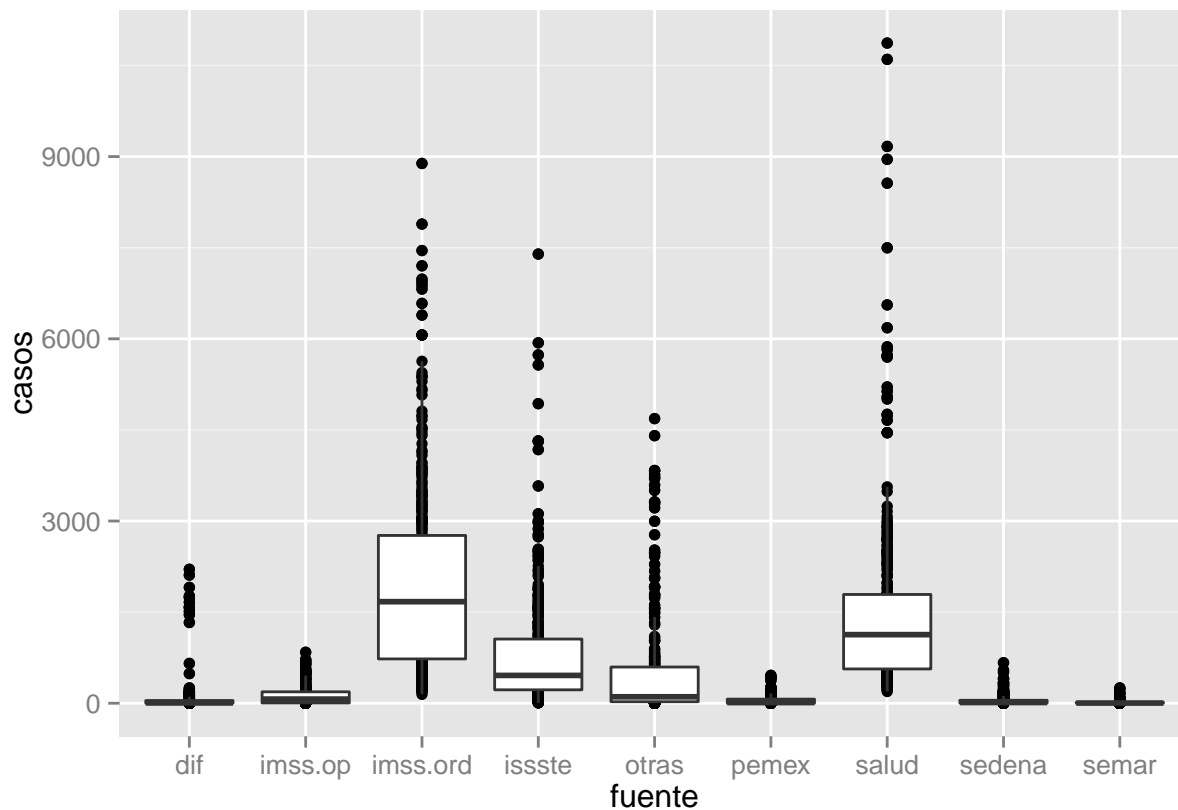
```
## [1] 1343917
```

Desviacion estandar.

```
## [1] 1159.274
```

```
## pdf
## 2
```

```
## pdf
## 2
```



La variacion es menor, asi como la media, en comparacion con el general. Sin embargo, el comportamiento de los casos en hombres en cada una de las instituciones es muy similar al general, es decir, los la mayoría de los datos se encuentran cerca de la media pero con gran cantidad de casos atipicos.

En mujeres:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   13.0   175.0   957.9 1074.0 17420.0
```

Varianza.

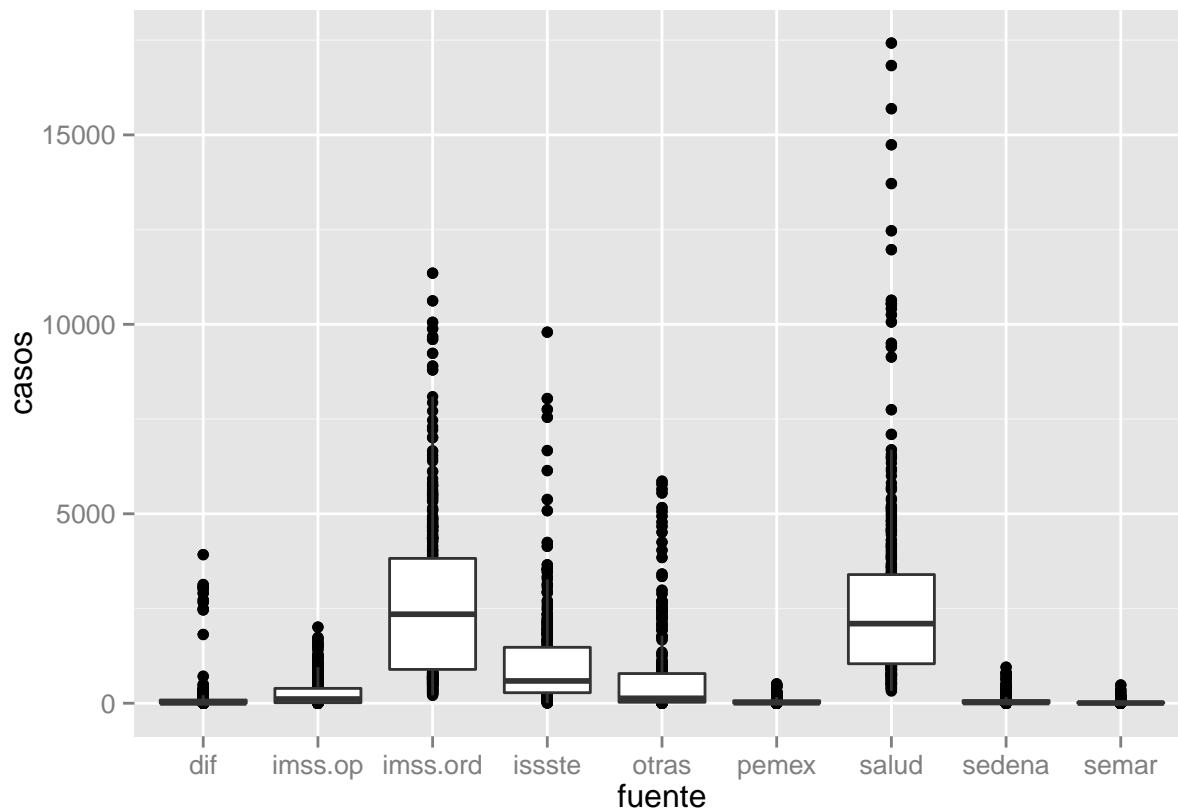
```
## [1] 2983395
```

Desviacion estandar.

```
## [1] 1727.251
```

```
## pdf
##      2
```

```
## pdf
##      2
```



Por ultimo, se presenta el promedio de casos anual en cada una de las instituciones:

```
##      Fuente      General
## 2 imss.ord 4696.65909
## 1  salud 4290.26136
## 3  issste 1864.26705
## 9  otras 1167.86916
## 4 imss.op 406.14826
## 5  dif 277.38168
## 7  sedena 109.68376
## 6  pemex 106.48230
## 8  semar 38.82031
```

## Analisis por Mes

El mayor numero de casos reportados ha sido en el mes de:

```
##      estado  genero fecha mes casos
## 7663 Mexico General 2011 ago 6667

##      estado  genero fecha mes casos
## 8015 Mexico Hombres 2011 ago 2858

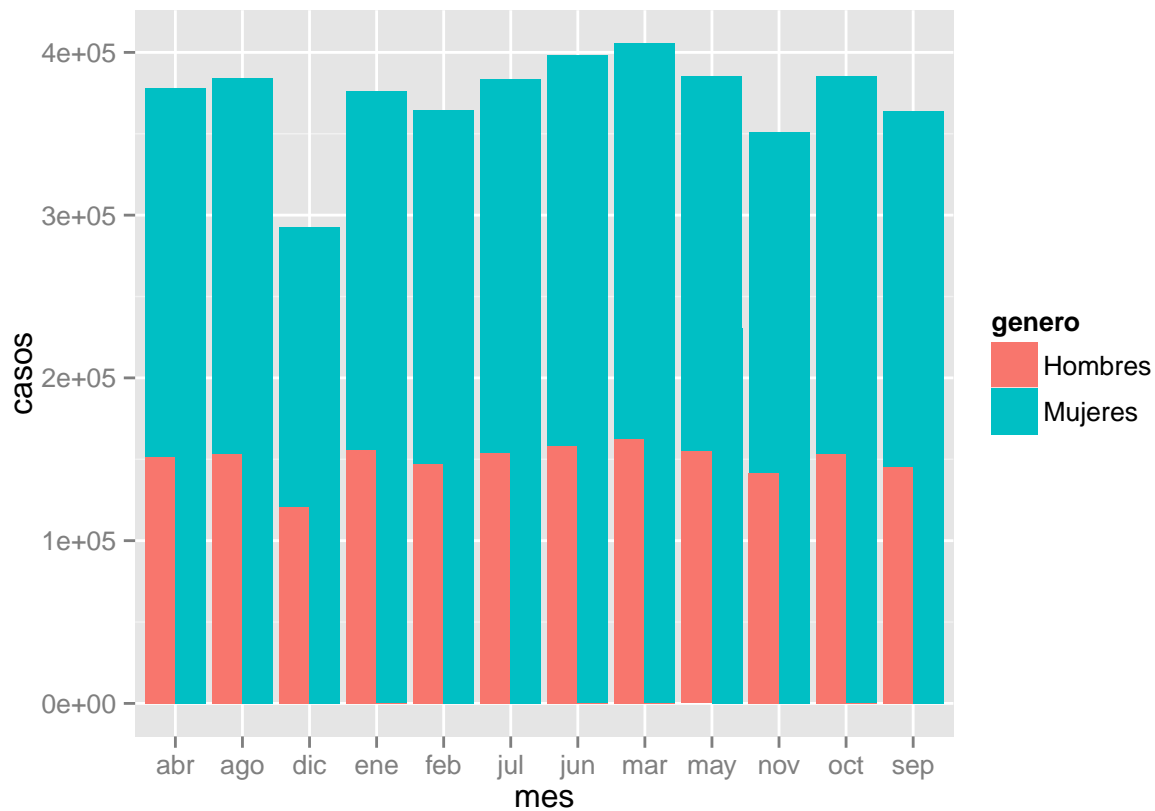
##      estado  genero fecha mes casos
## 8367 Mexico Mujeres 2011 ago 3809
```

El porcentaje reportes de diabetes se distribuye por cada mes de la siguiente manera:

```
## Mes General
## 3 mar 9.077486
## 6 jun 8.914162
## 5 may 8.629834
## 10 oct 8.626947
## 8 ago 8.600070
## 7 jul 8.583531
## 4 abr 8.461184
## 1 ene 8.417612
## 2 feb 8.152485
## 9 sep 8.143287
## 11 nov 7.849157
## 12 dic 6.544246
```

```
## pdf
## 2
```

```
## pdf
## 2
```



A nivel general, los casos tienen el siguiente comportamiento mensualmente:

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 68 417 902 1058 1282 6667
```

Varianza.

```
## [1] 821022.5
```

Desviacion estandar.

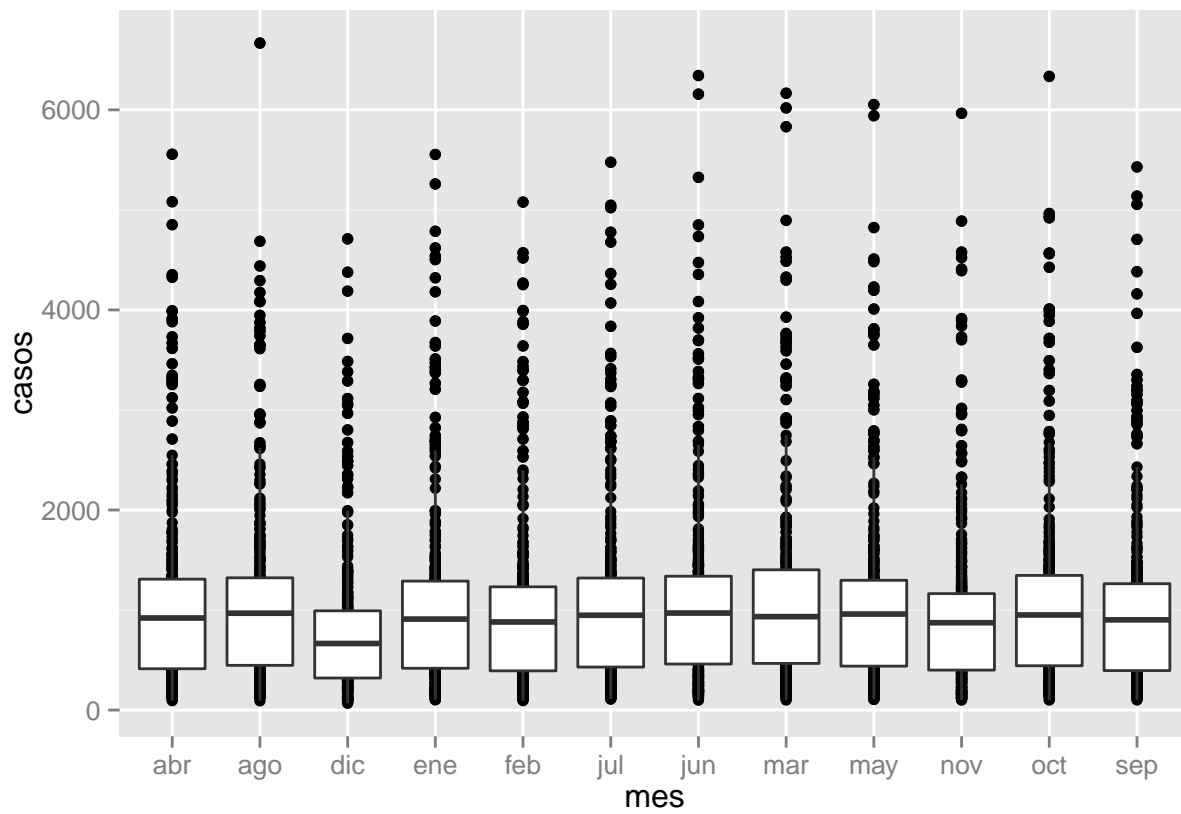
```
## [1] 906.1029
```

```
## pdf
```

```
## 2
```

```
## pdf
```

```
## 2
```



En hombres, el comportamiento es el siguiente:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      29.0  169.8   354.0   425.4  513.0  2858.0
```

Varianza.

```
## [1] 137395.6
```

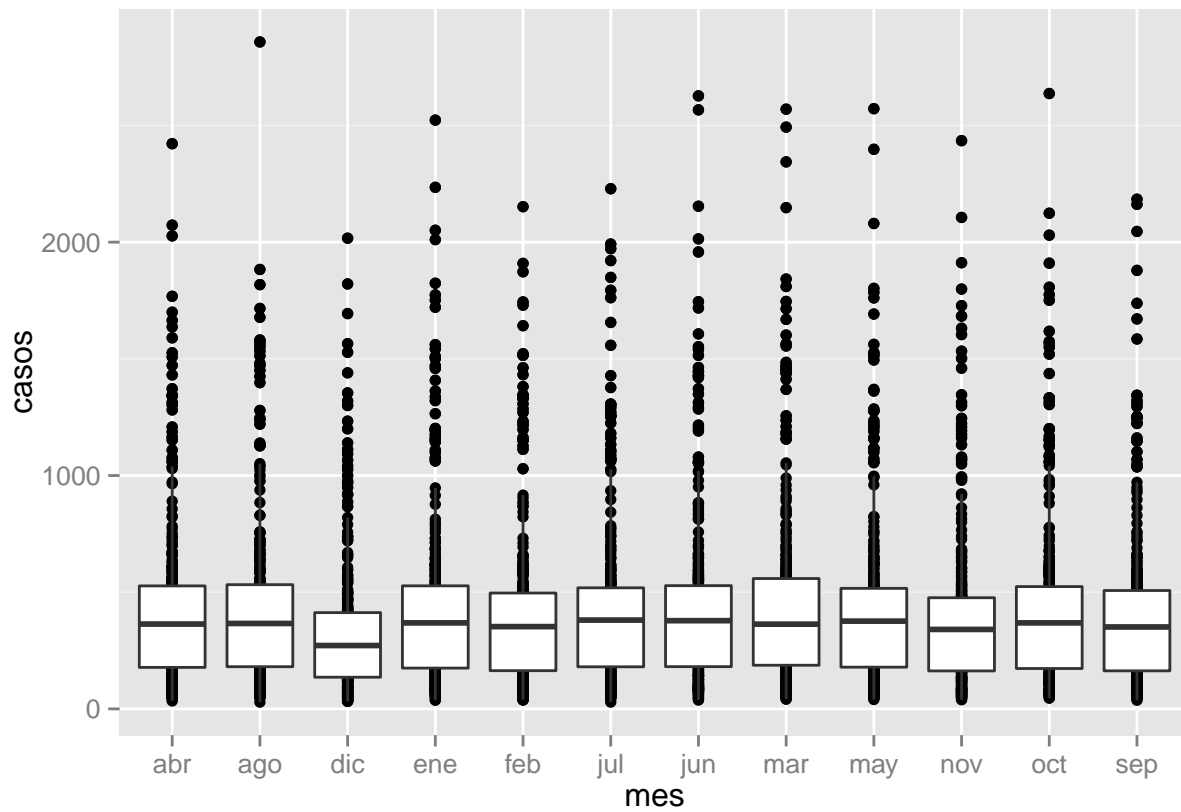
Desviacion estandar.

```
## [1] 370.6691
```



```
## pdf
## 2
```

```
## pdf
## 2
```



En mujeres:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      36.0  248.0   540.0   632.5  769.2  3809.0
```

Varianza.

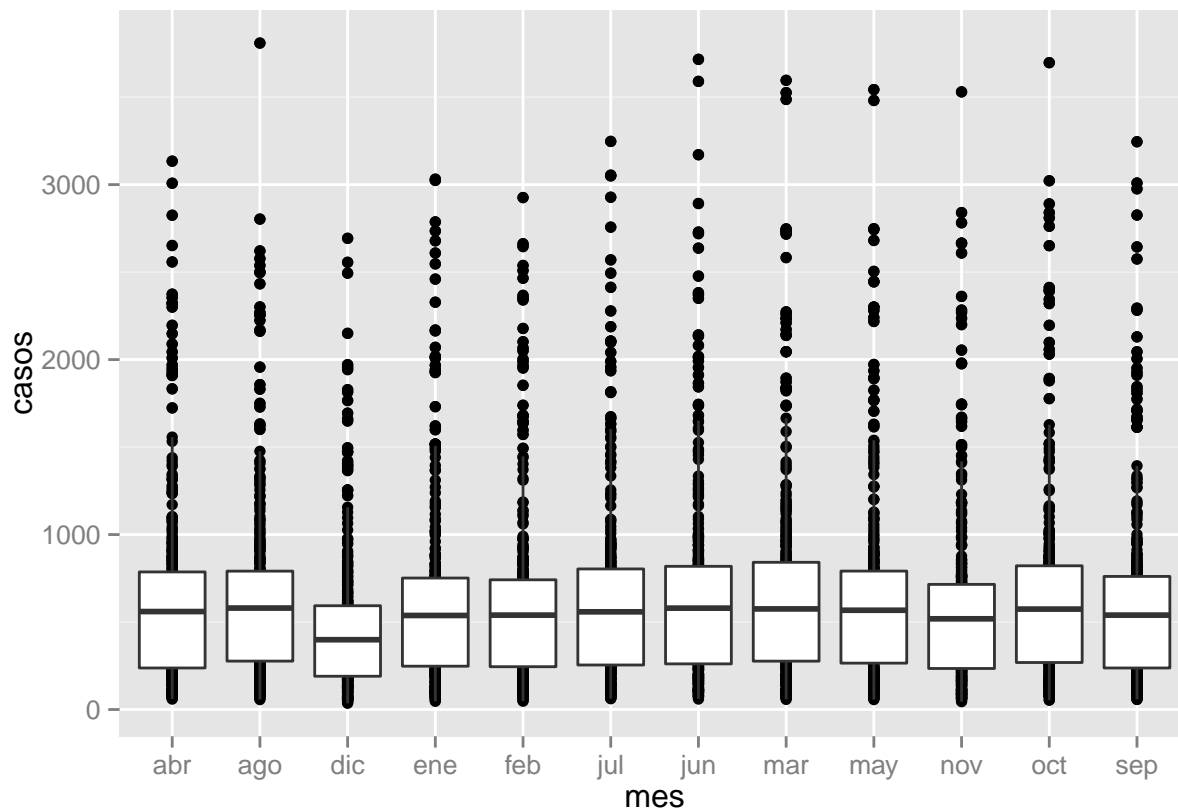
```
## [1] 289720.9
```

Desviacion estandar.

```
## [1] 538.2572
```

```
## pdf
## 2
```

```
## pdf
## 2
```



Por ultimo, se presenta el promedio mensual:

```
## Mes General
## 3 mar 1152.3324
## 6 jun 1131.5994
## 5 may 1095.5057
## 10 oct 1095.1392
## 8 ago 1091.7273
## 7 jul 1089.6278
## 4 abr 1074.0966
## 1 ene 1068.5653
## 2 feb 1034.9091
## 9 sep 1033.7415
## 11 nov 996.4034
## 12 dic 830.7528
```

## Analisis por Edad

El mayor numero de casos reportados por grupo de edad es::

```
## estado genero fecha edad casos
## 8463 Mexico General 2012 50a59 17236

## estado genero fecha edad casos
## 8847 Mexico Hombres 2013 50a59 7087
```

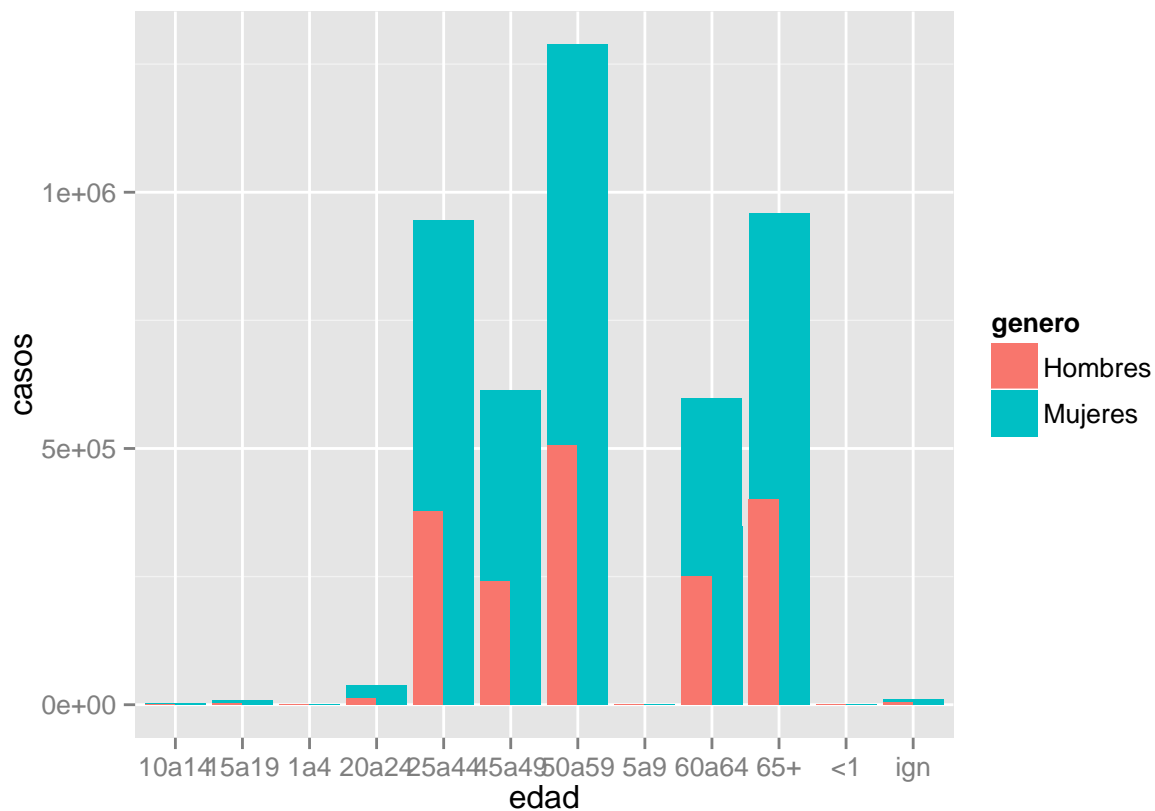
```
##      estado  genero fecha  edad casos
## 9071 Mexico Mujeres  2009 50a59 10290
```

El porcentaje reportes de diabetes se distribuye por cada grupo de edad de la siguiente manera:

```
##      Edad      General
## 9  50a59 28.82836451
## 11 65+ 21.45237174
## 7  25a44 21.18171733
## 8  45a49 13.75691546
## 10 60a64 13.39168643
## 6  20a24 0.85050025
## 12 ign 0.23106555
## 5  15a19 0.18583713
## 4  10a14 0.06203523
## 3   5a9 0.02607180
## 2   1a4 0.01788100
## 1    <1 0.01555356
```

```
## pdf
## 2
```

```
## pdf
## 2
```



A nivel general, los casos tienen el siguiente comportamiento por grupo de edad:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0      3.0     60.5  1082.0  1618.0 17240.0
```

Varianza.

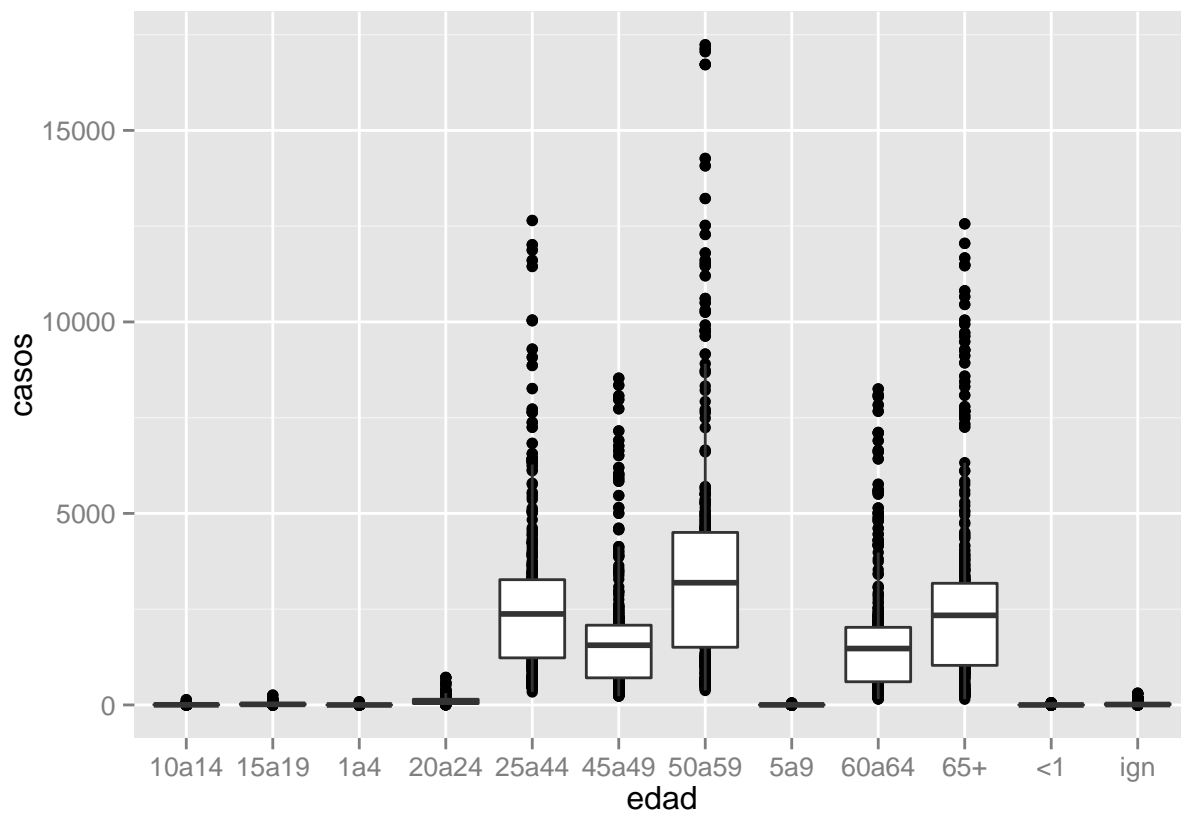
```
## [1] 3756380
```

Desviacion estandar.

```
## [1] 1938.138
```

```
## pdf
##      2
```

```
## pdf
##      2
```



En hombres, el comportamiento es el siguiente:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0      1.0     22.0   435.3   647.2  7087.0
```

Varianza.

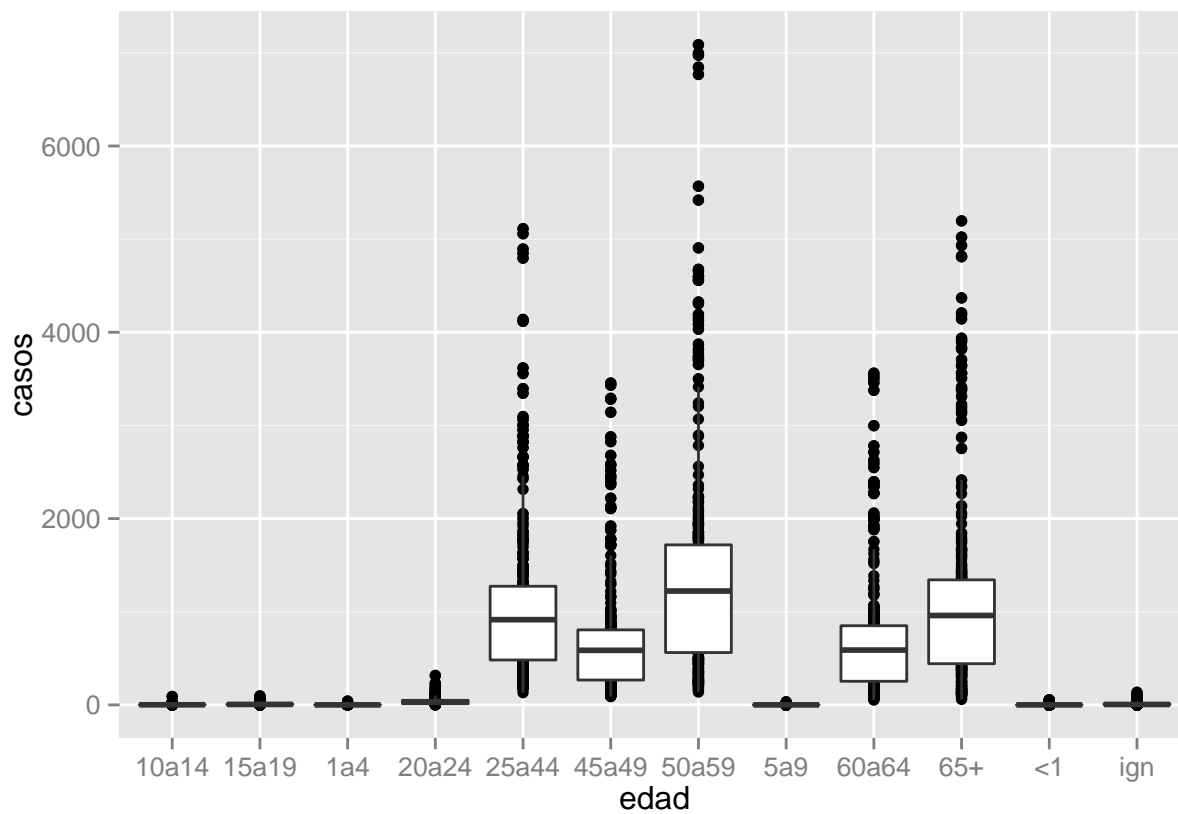
```
## [1] 616246.8
```

Desviacion estandar.

```
## [1] 785.0139
```

```
## pdf  
## 2
```

```
## pdf  
## 2
```



En mujeres:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      0.0      1.0    38.0   647.2   963.5  10290.0
```

Varianza.

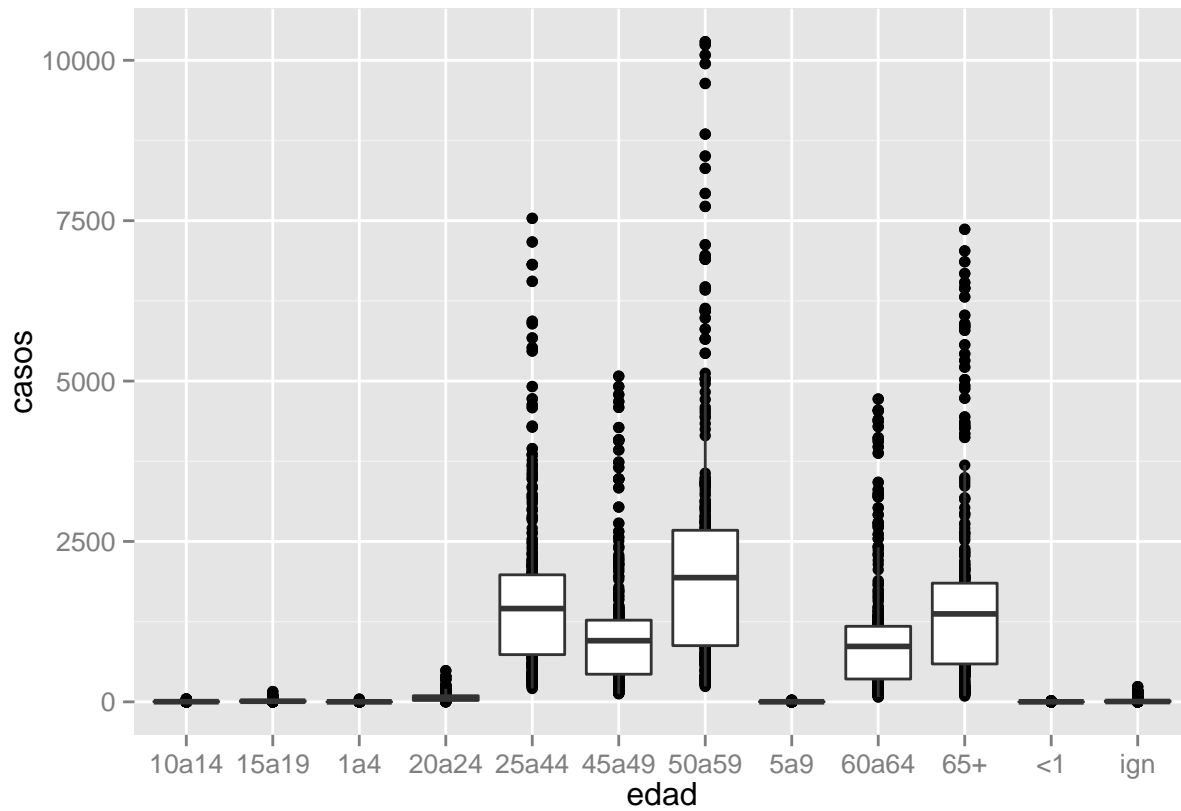
```
## [1] 1337848
```

Desviacion estandar.

```
## [1] 1156.654
```

```
## pdf  
## 2
```

```
## pdf
## 2
```



Por ultimo, se presenta el promedio por grupo de edad:

```
##      Edad      General
## 9  50a59 3659.588068
## 11 65+ 2723.250000
## 7  25a44 2688.892045
## 8  45a49 1746.357955
## 10 60a64 1699.994318
## 6  20a24 107.965909
## 12 ign 29.332386
## 5  15a19 23.590909
## 4  10a14 7.875000
## 3   5a9 3.640625
## 2   1a4 2.496875
## 1    <1 2.171875
```

**VII. Conclusiones** A nivel general, podemos concluir que, existe una tendencia creciente en el numero de casos reportados de diabetes, tanto para hombres como para mujeres. Aunque, el menor numero de casos en el periodo 2003-2013 se haya registrade en el ultimo anio (para mujeres).

Claramente podemos concluir que, las mujeres tienen una mayor propencion la diabetes melitus tipo II tanto a nivel general como en cada una de las categorias, tipo de fuente, mes y grupo de edad.

Siendo el estado de Mexico, el IMSS.ORD, el mes de marzo y el grupo de edad de 50 a 59 anios las categorias con mayor numero de casos de diabetes, al igual que el mayor promedio. Sin embargo, seria dificil dar una aproximacion basandonos solo en los promedios ya que los datos tienen una variacion muy alta, lo cual se representa en los casos atipicos a lo largo del analisis.

Una de las recomendaciones, para analisis posteriores, seria comparar los resultados actuales con los resultados de la misma base de datos pero eliminando los casos atipicos.

Tambien seria de gran interes conocer la relacion existente entre los casos de diabetes de hombres y mujeres en cada una de las categorias.