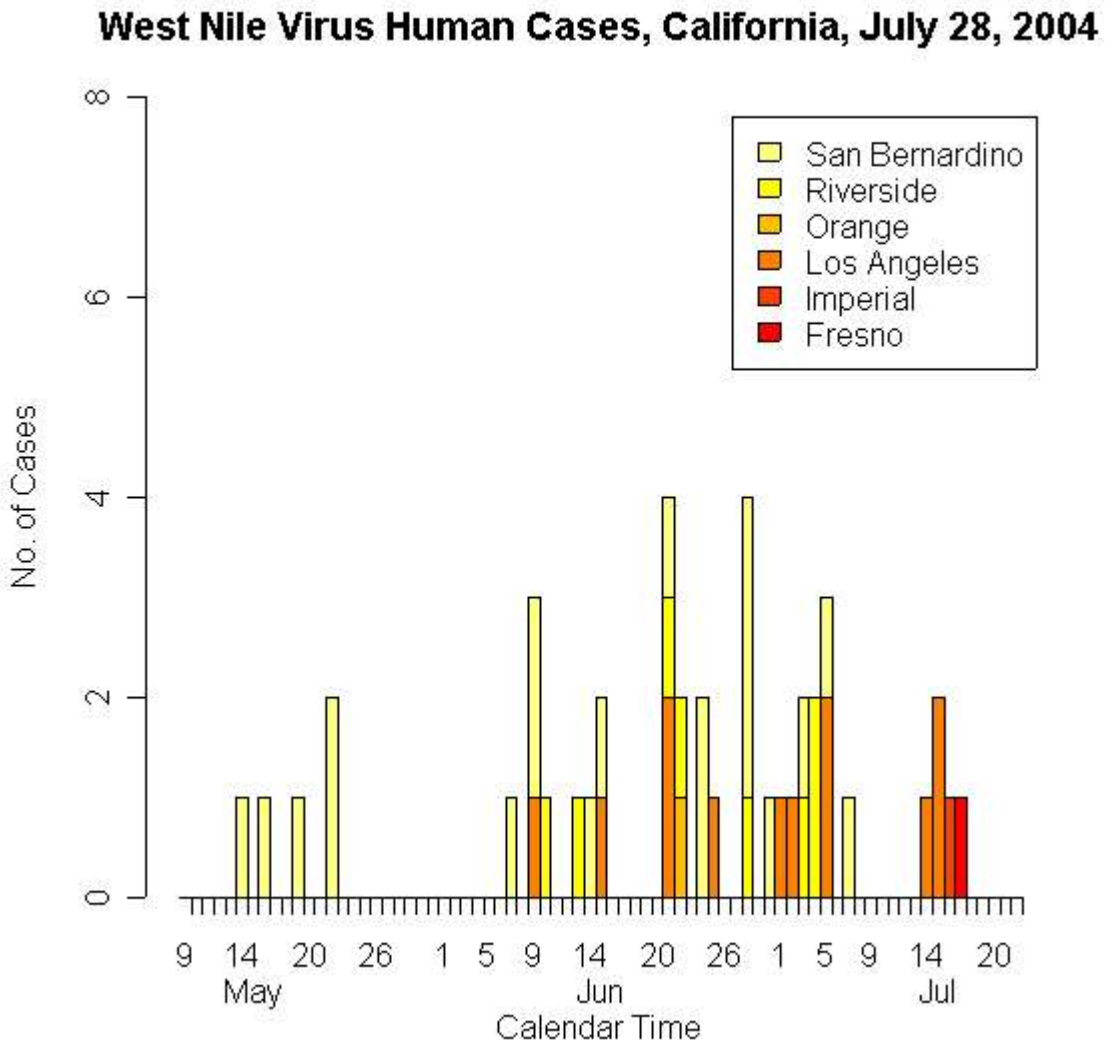
# Field Epidemiology Basics Computer Lab 2

*Example of Constructing an Epidemic Curve*
*Modified 2004-08-01*

## Epidemic curve

This document can be downloaded at
http://www.medepi.net/data/wnv/cidp-r-exercise2-example.pdf.



## Data

Data distributed by CDHS on July 28, 2004, and available at http://www.medepi.net/data/wnv/ca-wnv-2004-07-28.txt.

```
County, Age, Sex, Syndrome, Onset.Date, Date.Tested, Fatality
"San Bernardino",40,F,WNF,05/19/04,06/02/04,No
"San Bernardino",64,F,WNF,05/22/04,06/16/04,No
"San Bernardino",19,M,WNF,05/22/04,06/16/04,No
"San Bernardino",12,M,WNF,05/16/04,06/16/04,No
"San Bernardino",12,M,WNF,05/14/04,06/16/04,No
"San Bernardino",17,M,WNF,06/07/04,06/17/04,No
"San Bernardino",61,M,WNND,06/09/04,06/18/04,No
"San Bernardino",74,F,WNND,06/14/04,06/22/04,No
"Los Angeles",71,M,WNF,06/09/04,06/24/04,No
"Riverside",26,M,WNND,06/13/04,06/24/04,No
"Los Angeles",60,M,WNF,mid-June,06/25/04,No
"San Bernardino",84,F,WNND,mid-June,07/02/04,No
"San Bernardino",42,F,WNF,06/09/04,07/05/04,No
"San Bernardino",50,M,WNND,06/21/04,07/05/04,No
"Riverside",43,F,WNND,06/22/04,07/05/04,No
"Riverside",52,M,WNND,06/10/04,07/06/04,No
"San Bernardino",15,M,WNND,06/30/04,07/08/04,No
"San Bernardino",53,M,WNF,06/28/04,07/09/04,No
"San Bernardino",22,M,WNND,06/28/04,07/09/04,No
"San Bernardino",76,F,WNF,06/24/04,07/09/04,No
"Los Angeles",43,M,WNF,06/21/04,07/13/04,No
"Los Angeles",52,M,WNF,06/21/04,07/13/04,No
"San Bernardino",35,F,Unk,Unk,07/14/04,Unk
"San Bernardino",84,M,WNND,07/03/04,07/14/04,No
"Los Angeles",70,M,WNND,06/25/04,07/16/04,No
"Los Angeles",59,M,WNND,07/02/04,07/16/04,No
"Los Angeles",59,M,WNND,07/01/04,07/16/04,No
"San Bernardino",72,M,WNND,Unk,07/16/04,No
"Los Angeles",35,M,WNF,07/05/04,07/20/04,No
"San Bernardino",69,M,WNND,07/05/04,07/20/04,No
"San Bernardino",72,F,WNF,06/28/04,07/20/04,No
"Riverside",26,F,WNF,06/28/04,07/20/04,No
"Riverside",37,M,WNND,06/21/04,07/21/04,No
"Riverside",30,F,WNF,07/03/04,07/21/04,No
"Orange",57,M,WNND,06/22/04,07/21/04,Yes
"San Bernardino",52,M,Unk,Unk,07/22/04,Unk
"San Bernardino",59,M,Unk,Unk,07/22/04,Unk
"San Bernardino",35,F,WNF,07/07/04,07/22/04,No
"Riverside",56,M,WNND,07/04/04,07/22/04,No
"Riverside",46,M,WNF,07/04/04,07/22/04,No
"San Bernardino",45,M,Unk,Unk,07/23/04,Unk
"San Bernardino",24,M,WNF,06/24/04,07/23/04,No
"Imperial",55,F,WNF*,07/16/04,07/27/04,No
"San Bernardino",58,M,Unk,Unk,07/27/04,Unk
"San Bernardino",43,M,WNF,Unk,N/A,Unk
"San Bernardino",45,M,WNND,Unk,N/A,Unk
"Los Angeles",34,M,WNND,7/14/04,N/A,No
"Los Angeles",60,M,WNND,7/15/04,N/A,No
"Los Angeles",56,M,WNND,7/15/04,N/A,No
"Los Angeles",9,M,WNF,7/5/04,N/A,No
"Fresno",52,M,WNND,7/17/04,7/28/04,No
```

## R code in text editor

This code is available at http://www.medepi.net/data/wnv/ca-wnv-rjob.txt.

```
##READ DATA
wd <- read.table("http://www.medepi.net/data/wnv/ca-wnv-2004-07-28.txt",
      header=T, sep = ",", as.is = T, na.strings = "Unk")

##REVIEW AND CLEAN DATA
names(wd)
table(wd$County)
table(wd$Age)
hist(wd$Age, col = rgb(141, 160, 203, max = 250) )
summary(wd$Age)
table(wd$Sex)
table(wd$Syndrome)
wd$Syndrome[wd$Syndrome=="WNF*"] <- "WNF"
table(wd$Syndrome)
table(wd$Onset.Date)
wd$Onset.Date[wd$Onset.Date=="mid-June"] <- "06/15/04"
table(wd$Onset.Date)
table(wd$Date.Tested)
wd$Date.Tested[wd$Date.Tested=="N/A"] <- NA
table(wd$Date.Tested)
table(wd$Fatality)

##CREAT EPI CURVE
##First in R: Packages > Load Packages from CRAN > load chron

yy <- 5 #pad beginning and end of x axis
od <- chron(wd$Onset.Date)
caldates <- seq(min(od, na.rm=T) - yy, max(od, na.rm=T) + yy, by = 1)
caldays <- days(caldates)
odf <- factor(od, levels = caldates)
epidat <- table(wd$County, odf)
xv <- barplot(epidat,
              space = 0,
              axisnames = F,
              legend.text = T,
              axes = F,
              ylim = c(0, max(colSums(epidat))*2),
              xlab = "Calendar Time",
              ylab = "No. of Cases",
              main = "West Nile Virus Human Cases, California, July 28, 2004",)
axis(1, at = xv, labels = caldays, tick =F)
axis(1, at = xv+.5, labels = F, tick =T)
axis(2)
mtext(c("May","June","July"), side = 1, line = 2, at = xv[caldays==15])
```

# R code executed in R

```
> ##READ DATA
> wd <- read.table("http://www.medepi.net/data/wnv/ca-wnv-2004-07-28.txt",
     header=T, sep = ",", as.is = T, na.strings = "Unk")
>
> ##REVIEW AND CLEAN DATA
> names(wd)
[1] "County"      "Age"         "Sex"         "Syndrome"    "Onset.Date"
[6] "Date.Tested" "Fatality"
> table(wd$County)

        Fresno       Imperial    Los Angeles         Orange      Riverside
             1              1             12              1              8
San Bernardino
            28
> table(wd$Age)

 9 12 15 17 19 22 24 26 30 34 35 37 40 42 43 45 46 50 52 53 55 56 57 58 59 60
 1  2  1  1  1  1  1  2  1  1  3  1  1  1  3  2  1  1  4  1  1  2  1  1  3  2
61 64 69 70 71 72 74 76 84
 1  1  1  1  1  2  1  1  2
> hist(wd$Age, col = rgb(141, 160, 203, max = 250) )
> summary(wd$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   9.00   35.00   52.00   47.49   59.50   84.00
> table(wd$Sex)

 F  M
13 38
> table(wd$Syndrome)

 WNF WNF* WNND
  22    1   23
> wd$Syndrome[wd$Syndrome=="WNF*"] <- "WNF"
> table(wd$Syndrome)

 WNF WNND
  23   23
> table(wd$Onset.Date)

05/14/04 05/16/04 05/19/04 05/22/04 06/07/04 06/09/04 06/10/04 06/13/04
       1        1        1        2        1        3        1        1
06/14/04 06/21/04 06/22/04 06/24/04 06/25/04 06/28/04 06/30/04 07/01/04
       1        4        2        2        1        4        1        1
07/02/04 07/03/04 07/04/04 07/05/04 07/07/04 07/16/04  7/14/04  7/15/04
       1        2        2        2        1        1        1        2
 7/17/04   7/5/04 mid-June
       1        1        2
> wd$Onset.Date[wd$Onset.Date=="mid-June"] <- "06/15/04"
```

```
> table(wd$Onset.Date)

05/14/04 05/16/04 05/19/04 05/22/04 06/07/04 06/09/04 06/10/04 06/13/04
       1        1        1        2        1        3        1        1
06/14/04 06/15/04 06/21/04 06/22/04 06/24/04 06/25/04 06/28/04 06/30/04
       1        2        4        2        2        1        4        1
07/01/04 07/02/04 07/03/04 07/04/04 07/05/04 07/07/04 07/16/04  7/14/04
       1        1        2        2        2        1        1        1
 7/15/04  7/17/04   7/5/04
       2        1        1
> table(wd$Date.Tested)

06/02/04 06/16/04 06/17/04 06/18/04 06/22/04 06/24/04 06/25/04 07/02/04
       1        4        1        1        1        2        1        1
07/05/04 07/06/04 07/08/04 07/09/04 07/13/04 07/14/04 07/16/04 07/20/04
       3        1        1        3        2        2        4        4
07/21/04 07/22/04 07/23/04 07/27/04  7/28/04      N/A
       3        5        2        2        1        6
> wd$Date.Tested[wd$Date.Tested=="N/A"] <- NA
> table(wd$Date.Tested)

06/02/04 06/16/04 06/17/04 06/18/04 06/22/04 06/24/04 06/25/04 07/02/04
       1        4        1        1        1        2        1        1
07/05/04 07/06/04 07/08/04 07/09/04 07/13/04 07/14/04 07/16/04 07/20/04
       3        1        1        3        2        2        4        4
07/21/04 07/22/04 07/23/04 07/27/04  7/28/04
       3        5        2        2        1
> table(wd$Fatality)

 No Yes
 43   1
>
> ##CREAT EPI CURVE
> ##First in R: Packages > Load Packages from CRAN > load chron
>
> yy <- 5 #pad beginning and end of x axis
> od <- chron(wd$Onset.Date)
>
> caldates <- seq(min(od, na.rm=T) - yy, max(od, na.rm=T) + yy, by = 1)
> caldays <- days(caldates)
> odf <- factor(od, levels = caldates)
> epidat <- table(wd$County, odf)
> xv <- barplot(epidat,
+               space = 0,
+               axisnames = F,
+               legend.text = T,
+               axes = F,
+               ylim = c(0, max(colSums(epidat))*2),
+               xlab = "Calendar Time",
```

```
+               ylab = "No. of Cases",
+               main = "West Nile Virus Human Cases, California, July 28, 2004",)
> axis(1, at = xv, labels = caldays, tick =F)
> axis(1, at = xv+.5, labels = F, tick =T)
> axis(2)
> mtext(c("May","June","July"), side = 1, line = 2, at = xv[caldays==15])
```