ACME_Data_Analysis.R

puj83

Mon Jul 29 20:32:09 2019

```r
# Brief background- Urinalyses are an important part of a patient work up in
order to assess kidney function. UAs consist of physical, chemical and micros
copic analyses of the urine.
# More specific information about UAs and their interpretation can be found h
ere (http://www.eclinpath.com/urinalysis/).
# While dipstick tests are commonly used as a sensitive measure to detect pro
teinuria (the presence of protein in the urine), the urine protein:creatinine
ratio is recognized as a more accurate assessment of proteinuria.
# UPCs are commonly used as a follow up test when protein is detected in UA r
esults via dipstick, as it requires a separate workflow in the laboratory and
is run on a chemistry analyzer.
# Additional information can be found here (http://www.eclinpath.com/urinalys
is/chemical-constituents/).
#
# ACME is a corporate account that has been running UPCs (urine protein creat
inine ratios) on most of their UAs (urinalyses) they send in to the MedX refe
rence lab.
# Their pricing is changing and they will no longer be receiving UPCs for fre
e.
# Please use ACME's data from the last year to help the professional service
veterinarian and corporate accounts manager communicate to ACME's veterinaria
ns
# that they have been running unnecessary UPCs and that MedX's Urinalysis wit
h Reflex UPCs offering might be a good fit for them and help them reduce unne
cessary testing when UPCs are not appropriate.
# Please use the indicated parameters found below in the MedX Test Directory
for your analyses.
# MedX Test Directory info: Urinalysis with Reflex UPC (If Indicated) (2326)
# If the urinalysis is positive for protein and the sediment is not active, a
urine protein:creatinine (UPC) ratio is automatically performed.
# If the urinalysis is negative for protein, or if there is an active sedimen
t (white blood cells ???6/hpf, red blood cells ??? 100/hpf, color is red or p
ink, and/or bacteria are seen),
# the UPC ratio will not be performed. The test price is the same whether or
not a UPC ratio is performed.




# install.packages("ggplot2")
# install.packages("stringr")
# install.packages("dplyr")
```

```r
# install.packages("plyr")

library(stringr)
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(plyr)

## --------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, th
en dplyr:
## library(plyr); library(dplyr)

## --------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

ACME_Data<-read.csv(file = "C:/Users/puj83/OneDrive/CV/Cases/MedX/ACME_Data.c
sv", header = T, sep = ",")
ACME_Data$Result.Text.URINALYSIS.BACTERIA <- as.character(ACME_Data$Result.Te
xt.URINALYSIS.BACTERIA)
ACME_Data$Result.Text.URINALYSIS.BACTERIA[ACME_Data$Result.Text.URINALYSIS.BA
CTERIA==""] <- "NA"
ACME_Data$Result.Text.URINALYSIS.BACTERIA <- as.factor(ACME_Data$Result.Text.
URINALYSIS.BACTERIA)

# 1.    How many urinalyses were run?

ACME_Data$UA_Count <- ifelse(grepl("NA", ACME_Data$Result.Text.URINALYSIS.BAC
TERIA), "1", "0")
ACME_Data$UA_Count<-as.numeric(as.character(ACME_Data$UA_Count))

dfl <- ddply(ACME_Data, .(UA_Count), summarize, y=length(UA_Count))
```
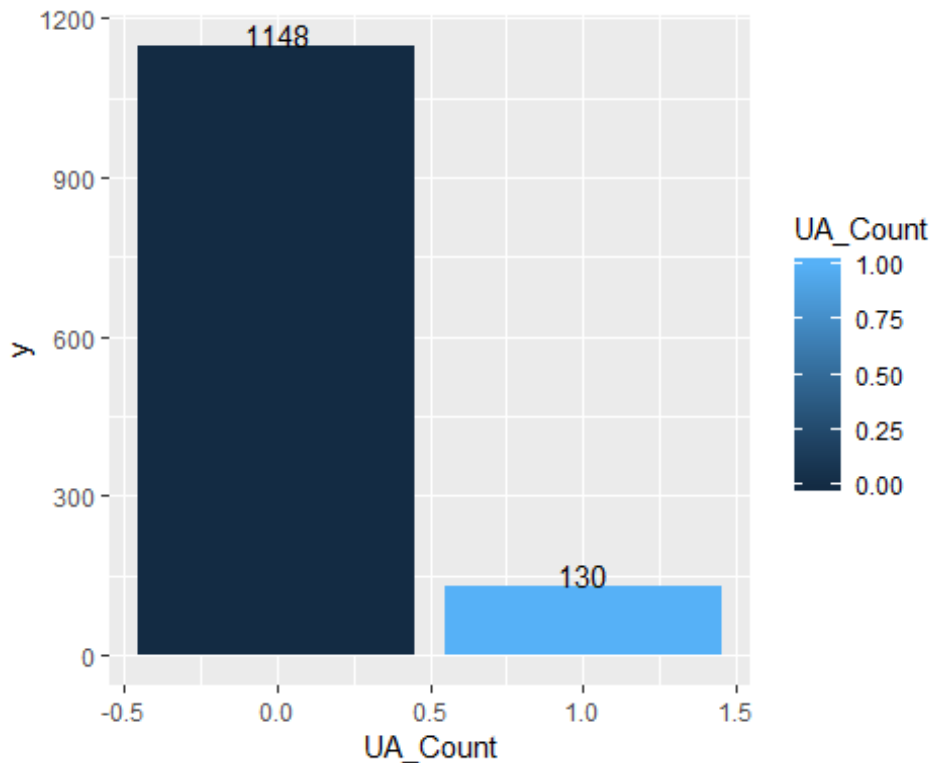
```
dfl$y<-as.numeric(as.character(dfl$y))

dfl<-as.data.frame(dfl)

ggplot(dfl, aes(UA_Count, y=y, fill=UA_Count)) + geom_bar(stat="identity") +
geom_text(aes(label=y), vjust=0)
```



```
# 2.    How many UAs were positive for protein?
```

```
ACME_Data$Result.Text.URINALYSIS.PROTEIN<-as.factor(ACME_Data$Result.Text.URI
NALYSIS.PROTEIN)
levels(ACME_Data$Result.Text.URINALYSIS.PROTEIN)

## [1] ""
## [2] "1+ (100-200 mg/dL)"
## [3] "2+ (200-300 mg/dL)"
## [4] "3+ (300-500 mg/dL)"
## [5] "4+ (>500 mg/dL)"
## [6] "DNR"
## [7] "INSUFFICIENT SAMPLE FOR COMPLETE ANALYSIS."
## [8] "NEGATIVE"
## [9] "TRACE"

ACME_Data$Result.Text.URINALYSIS.PROTEIN<-as.character(ACME_Data$Result.Text.
URINALYSIS.PROTEIN)
ACME_Data$UAs_Positive_Count <- ifelse(grepl("1+|2+|3+|4+", ACME_Data$Result.
Text.URINALYSIS.PROTEIN), "1", "0")
```

```
ACME_Data$UAs_Positive_Count<-as.numeric(as.character(ACME_Data$UAs_Positive_
Count))

df2 <- ddply(ACME_Data, .(UAs_Positive_Count), summarize, y=length(UAs_Positi
ve_Count))
df2$y<-as.numeric(as.character(df2$y))
df2<-as.data.frame(df2)

ggplot(df2, aes(UAs_Positive_Count, y=y, fill=UAs_Positive_Count)) + geom_bar
(stat="identity") +
  geom_text(aes(label=y), vjust=0)
```
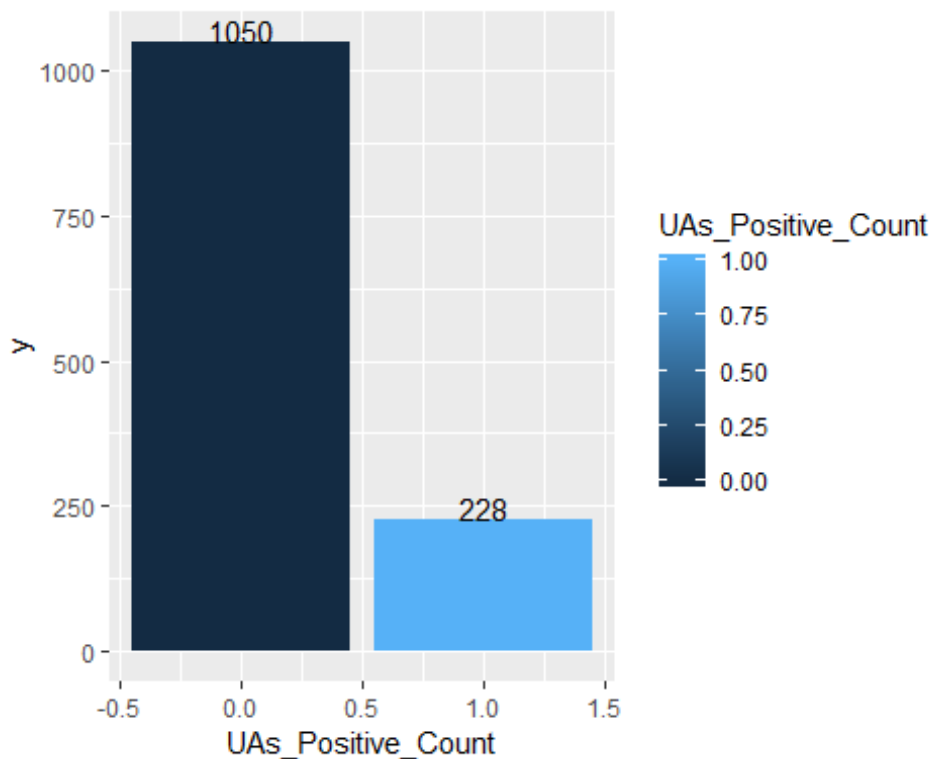


```
# 3.    How many UAs were positive for protein and had no active sediment? (N
umber of UPCs which would have been run IF ACME used the Reflex UPC test offe
ring)
```

```
ACME_Data$UAs_Positive_Count<-as.factor(ACME_Data$UAs_Positive_Count)
levels(ACME_Data$UAs_Positive_Count)

## [1] "0" "1"

ACME_Data$Result.Text.URINALYSIS.BACTERIA<-as.factor(ACME_Data$Result.Text.UR
INALYSIS.BACTERIA)
levels(ACME_Data$Result.Text.URINALYSIS.BACTERIA)

##  [1] "0"
##  [2] "DNR"
```

```
##  [3] "INSUFFICIENT SAMPLE FOR COMPLETE ANALYSIS."
##  [4] "MARKED (>40/HPF)"
##  [5] "MARKED >40/HPF, RODS AND COCCI PRESENT"
##  [6] "MARKED COCCI >40/HPF"
##  [7] "Marked Rods (>40/HPF)"
##  [8] "MARKED RODS >40/HPF"
##  [9] "MODERATE (9-40/HPF)"
## [10] "MODERATE 9-40/HPF, RODS AND COCCI PRESENT"
## [11] "MODERATE COCCI 9-40/HPF"
## [12] "MODERATE RODS 9-40/HPF"
## [13] "NA"
## [14] "NONE SEEN"
## [15] "RARE (<8/HPF)"
## [16] "RARE <9/HPF, RODS AND COCCI PRESENT"
## [17] "RARE COCCI <9/HPF"
## [18] "RARE RODS <9/HPF"
```

```r
ACME_Data$Result.Text.URINALYSIS.COLOR<-as.factor(ACME_Data$Result.Text.URINALYSIS.COLOR)
levels(ACME_Data$Result.Text.URINALYSIS.COLOR)
```

```
##  [1] ""            "AMBER"       "BROWN"       "COLORLESS"   "DARK YELLOW"
##  [6] "DNR"         "ORANGE"      "OTHER"       "PALE YELLOW" "PINK"
## [11] "RED"         "STRAW"       "Yellow"      "YELLOW"
```

```r
ACME_Data$Result.Text.URINALYSIS.RBC<-as.factor(ACME_Data$Result.Text.URINALYSIS.RBC)
levels(ACME_Data$Result.Text.URINALYSIS.RBC)
```

```
##  [1] ""
##  [2] ">100"
##  [3] "0-2"
##  [4] "10-Jun"
##  [5] "15-20"
##  [6] "15-Oct"
##  [7] "20-30"
##  [8] "30-50"
##  [9] "5-Feb"
## [10] "50-75"
## [11] "75-100"
## [12] "DNR"
## [13] "INSUFFICIENT SAMPLE FOR COMPLETE ANALYSIS."
## [14] "NONE SEEN"
```

```r
levels(ACME_Data$Result.Text.URINALYSIS.RBC)[levels(ACME_Data$Result.Text.URINALYSIS.RBC) == "10-Jun"]<-"6-10"
levels(ACME_Data$Result.Text.URINALYSIS.RBC)[levels(ACME_Data$Result.Text.URINALYSIS.RBC) == "15-Oct"]<-"10-15"
levels(ACME_Data$Result.Text.URINALYSIS.RBC)[levels(ACME_Data$Result.Text.URINALYSIS.RBC) == "5-Feb"]<-"2-5"
```

```r
ACME_Data$Result.Text.URINALYSIS.WBC<-as.factor(ACME_Data$Result.Text.URINALYSIS.WBC)
levels(ACME_Data$Result.Text.URINALYSIS.WBC)

## [1] ""
## [2] ">100"
## [3] "0-2"
## [4] "10-Jun"
## [5] "15-20"
## [6] "15-Oct"
## [7] "20-30"
## [8] "30-50"
## [9] "5-Feb"
## [10] "50-75"
## [11] "75-100"
## [12] "DNR"
## [13] "INSUFFICIENT SAMPLE FOR COMPLETE ANALYSIS."
## [14] "NONE SEEN"

levels(ACME_Data$Result.Text.URINALYSIS.WBC)[levels(ACME_Data$Result.Text.URINALYSIS.WBC) == "10-Jun"]<-"6-10"
levels(ACME_Data$Result.Text.URINALYSIS.WBC)[levels(ACME_Data$Result.Text.URINALYSIS.WBC) == "15-Oct"]<-"10-15"
levels(ACME_Data$Result.Text.URINALYSIS.WBC)[levels(ACME_Data$Result.Text.URINALYSIS.WBC) == "5-Feb"]<-"2-5"

ACME_Data$Result.Text.URINALYSIS.BACTERIA<-as.character(ACME_Data$Result.Text.URINALYSIS.BACTERIA)
ACME_Data$Bacteria_Q <- ifelse(grepl("0|NONE SEEN|RARE COCCI <9/HPF", ACME_Data$Result.Text.URINALYSIS.BACTERIA), "1", "0")
ACME_Data$Bacteria_Q<-as.numeric(as.character(ACME_Data$Bacteria_Q))

ACME_Data$Result.Text.URINALYSIS.COLOR<-as.character(ACME_Data$Result.Text.URINALYSIS.COLOR)
ACME_Data$Color_Q <- ifelse(grepl("AMBER|BROWN|COLORLESS|DARK YELLOW|ORANGE|OTHER|PALE YELLOW|STRAW|YELLOW", ACME_Data$Result.Text.URINALYSIS.COLOR), "1", "0")
ACME_Data$Color_Q<-as.numeric(as.character(ACME_Data$Color_Q))

ACME_Data$Result.Text.URINALYSIS.RBC<-as.character(ACME_Data$Result.Text.URINALYSIS.RBC)
ACME_Data$RBC_Q <- ifelse(grepl("0-2|6-10|15-20|10-15|20-30|30-50|2-5|50-75|75-100", ACME_Data$Result.Text.URINALYSIS.RBC), "1", "0")
ACME_Data$RBC_Q<-as.numeric(as.character(ACME_Data$RBC_Q))

ACME_Data$Result.Text.URINALYSIS.WBC<-as.character(ACME_Data$Result.Text.URINALYSIS.WBC)
ACME_Data$WBC_Q <- ifelse(grepl("0-2|2-5", ACME_Data$Result.Text.URINALYSIS.WBC), "1", "0")
ACME_Data$WBC_Q<-as.numeric(as.character(ACME_Data$WBC_Q))
```

```
ACME_Data$Positive_NoSediment<-with(ACME_Data, ifelse(ACME_Data$UAs_Positive_
Count == 1 &
                                        ACME_Data$Bacteria_Q ==
1 &
                                        ACME_Data$Color_Q == 1
&
                                        ACME_Data$RBC_Q == 1 &
                                        ACME_Data$WBC_Q == 1, 1
, 0))

df3 <- ddply(ACME_Data, .(Positive_NoSediment), summarize, y=length(Positive_
NoSediment))
df3$y<-as.numeric(as.character(df3$y))
df3<-as.data.frame(df3)

ggplot(df3, aes(Positive_NoSediment, y=y, fill=Positive_NoSediment)) + geom_b
ar(stat="identity") +
  geom_text(aes(label=y), vjust=0)
```
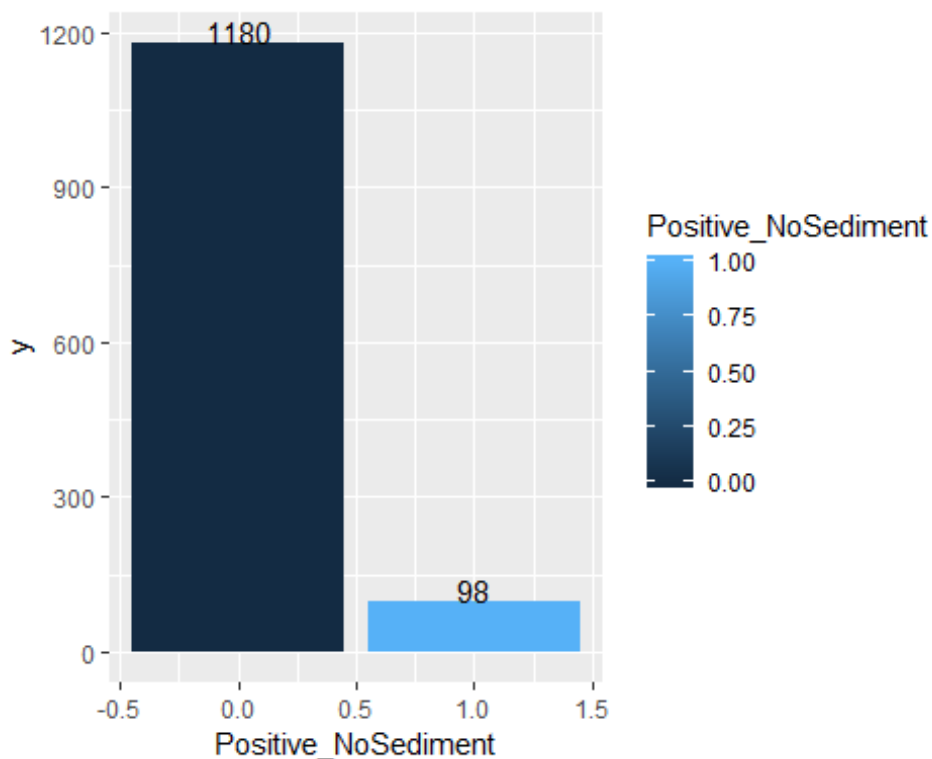


```
# 4.     Number of actual UPCs run:

#IF UPC was run cell would not be blank, if one cell is blank all three are

UPCRATIOBLANK<-is.na(ACME_Data$Result.Text.UPC.Ratio)
sum(UPCRATIOBLANK)
```

```
## [1] 0
```

```r
# 5.    How many UPCs were run that might have been unnecessary using the ref
lex UPC testing criteria?

ACME_Data$Result.Text.UPC.Ratio<-as.factor(ACME_Data$Result.Text.UPC.Ratio)
levels(ACME_Data$Result.Text.UPC.Ratio)
```

```
##  [1] ""
##  [2] "0"
##  [3] "0.1"
##  [4] "0.2"
##  [5] "0.3"
##  [6] "0.4"
##  [7] "0.5"
##  [8] "0.6"
##  [9] "0.7"
## [10] "0.8"
## [11] "0.9"
## [12] "1"
## [13] "1.1"
## [14] "1.2"
## [15] "1.3"
## [16] "1.4"
## [17] "1.5"
## [18] "1.6"
## [19] "1.7"
## [20] "1.8"
## [21] "1.9"
## [22] "10.4"
## [23] "10.6"
## [24] "11.2"
## [25] "17.8"
## [26] "2"
## [27] "2.1"
## [28] "2.2"
## [29] "2.3"
## [30] "2.4"
## [31] "2.5"
## [32] "2.6"
## [33] "2.7"
## [34] "2.8"
## [35] "2.9"
## [36] "3.1"
## [37] "3.2"
## [38] "3.3"
```

```
## [39] "3.4"
## [40] "3.5"
## [41] "3.6"
## [42] "3.9"
## [43] "4"
## [44] "4.1"
## [45] "4.2"
## [46] "4.3"
## [47] "4.5"
## [48] "4.6"
## [49] "4.7"
## [50] "4.8"
## [51] "5"
## [52] "5.2"
## [53] "5.4"
## [54] "5.5"
## [55] "5.6"
## [56] "5.7"
## [57] "5.8"
## [58] "6.1"
## [59] "6.2"
## [60] "6.4"
## [61] "6.6"
## [62] "6.9"
## [63] "7"
## [64] "7.2"
## [65] "7.3"
## [66] "7.4"
## [67] "7.7"
## [68] "8.6"
## [69] "8.8"
## [70] "9.4"
## [71] "9.8"
## [72] "INSUFFICIENT SAMPLE FOR COMPLETE ANALYSIS."
```

```r
ACME_Data$Result.Text.UPC.Ratio<-as.numeric(as.character(ACME_Data$Result.Text.UPC.Ratio))
```

```
## Warning: NAs introduced by coercion
```

```r
ACME_Data$Unnecessary_UPCs<-with(ACME_Data, ifelse(ACME_Data$UAs_Positive_Count == 0 &
                                                   ACME_Data$Result.Text.UPC.Ratio > 0, 1, 0))

df4 <- ddply(ACME_Data, .(Unnecessary_UPCs), summarize, y=length(Unnecessary_UPCs))
df4$y<-as.numeric(as.character(df4$y))
df4<-as.data.frame(df4)
```

```
ggplot(df4, aes(Unnecessary_UPCs, y=y, fill=Unnecessary_UPCs)) + geom_bar(sta
t="identity") +
  geom_text(aes(label=y), vjust=0)
```

Warning: Removed 1 rows containing missing values (position_stack).

Warning: Removed 1 rows containing missing values (geom_text).