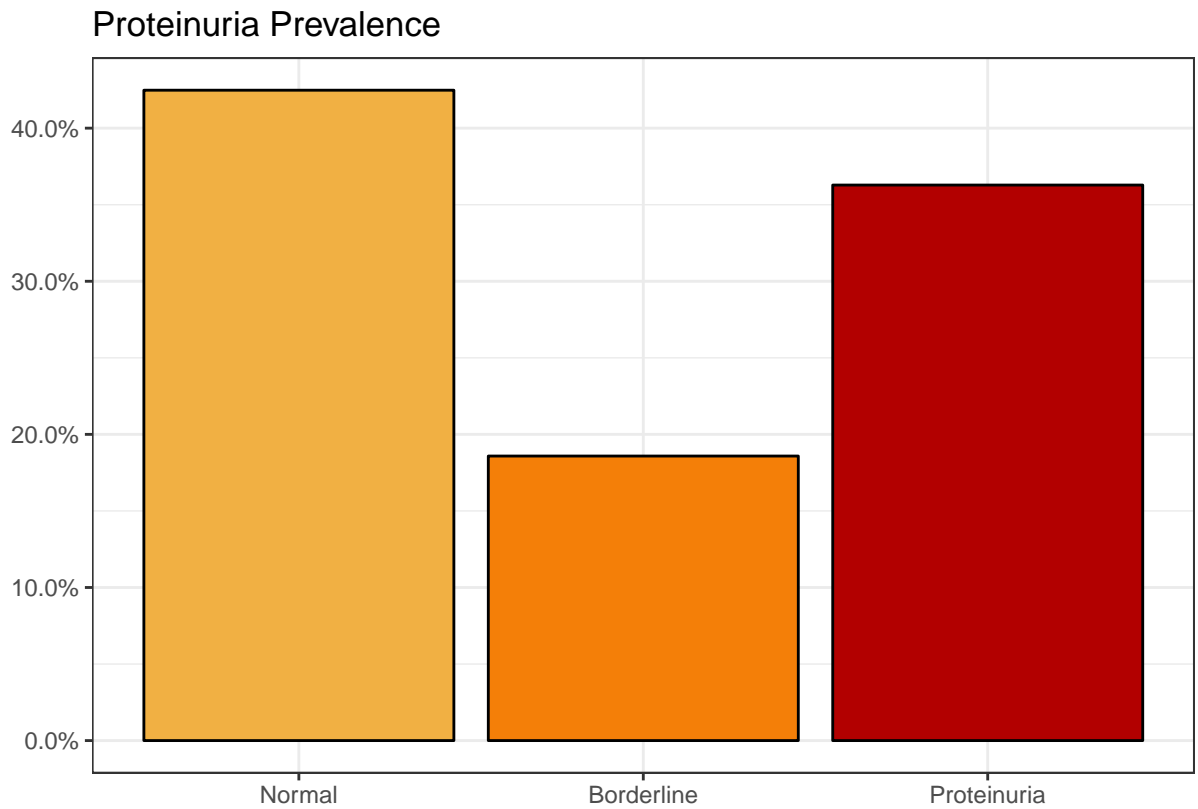


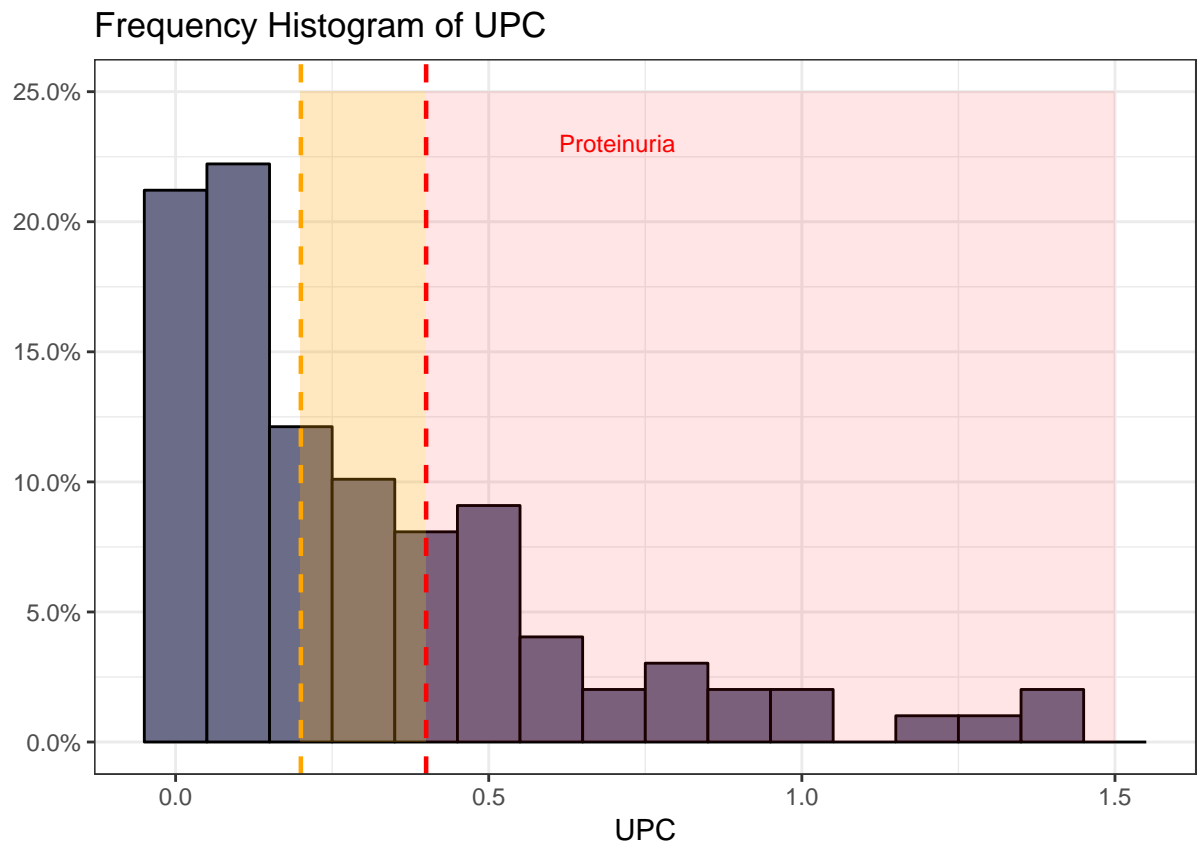
Plots

UPC



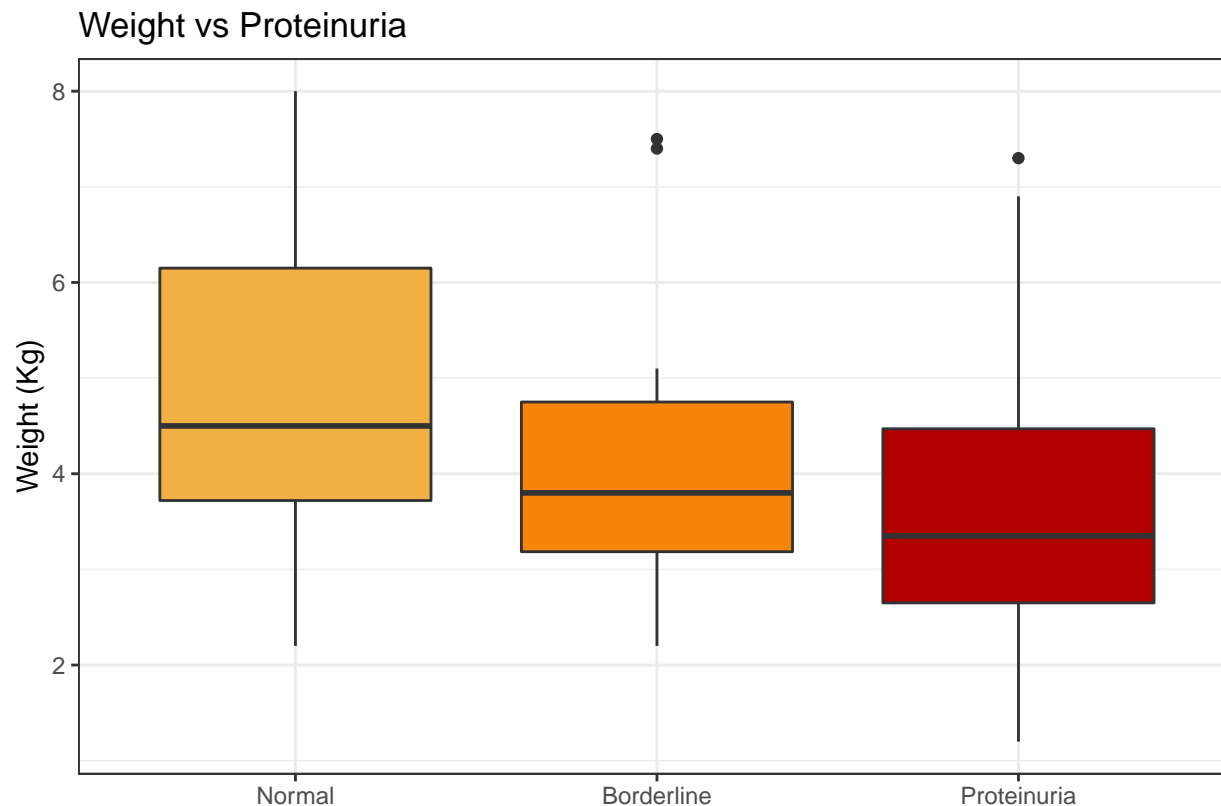
```
## # A tibble: 3 x 2
##   protein    percent
##   <fct>      <dbl>
## 1 Normal      42.5
## 2 Borderline  18.6
## 3 Proteinuria 36.3
```

Aquí només ensenyo la proporció dels nivells de proteinuria.



Aquest segon plot és una visió més detallada de l'anterior. Bàsicament tens la distribució dels nivells de UPCR. He pintat en taronja la zona **borderline** (entre 0.2 i 0.4) i en vermell la que he catalogat de **Proteinuria** (superior a 0.4).

Weight



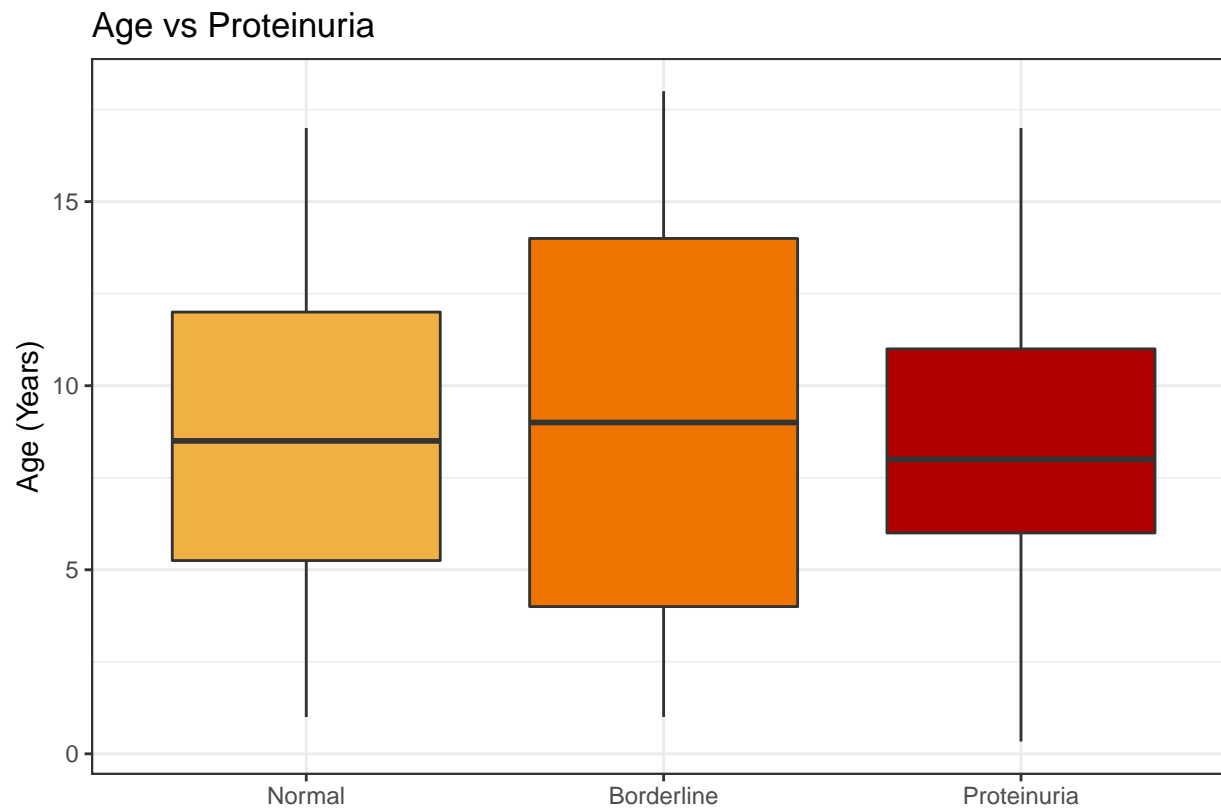
```
## # A tibble: 3 x 3
##   protein    weight_mean weight_sd
##   <fct>      <dbl>     <dbl>
## 1 Normal        4.83       1.60
## 2 Borderline    3.8       1.57
## 3 Proteinuria   3.3       1.59
```

Boxplot on es veu la mitja i la dispersió de les mesures de pes pels diferents nivells de proteinuria. Es veu que la proteinuria es relaciona amb menor pes.

```
## [1] "El resultat del test estadístic surt amb un p-value de 0.004. "
```

```
## [1] "El que ens diu que la diferencia és estadísticament significativa."
```

Age



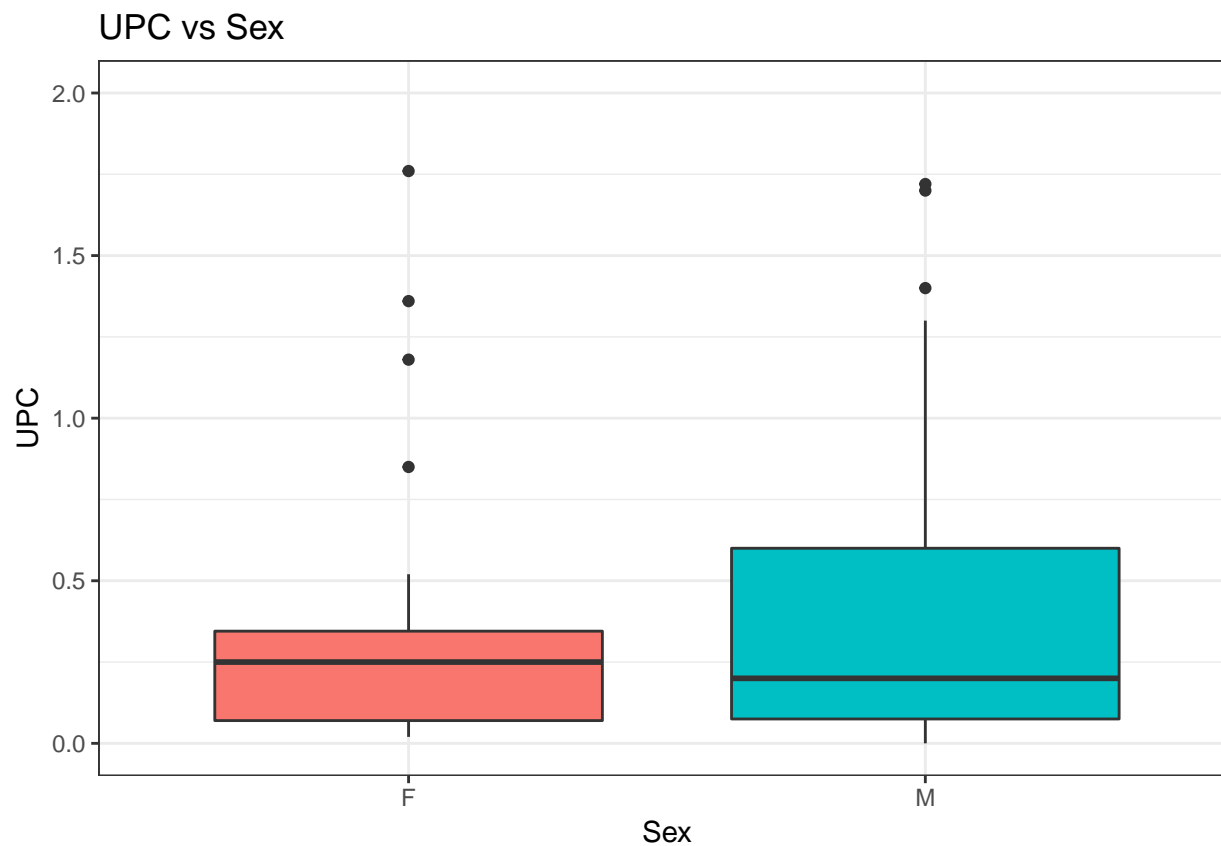
```
## # A tibble: 3 x 3
##   protein    age_mean age_sd
##   <fct>      <dbl>  <dbl>
## 1 Normal      8.45   4.07
## 2 Borderline  9.12   5.54
## 3 Proteinuria 8.62   4.41
```

Aquí no cal ni un test. És evident q la proteinuria no enten d'edats. Però el fem igual

```
## [1] "El resultat del test estadístic surt amb un p-value de 0.87. "
```

```
## [1] "Per tant la diferencia NO és estadísticament significativa."
```

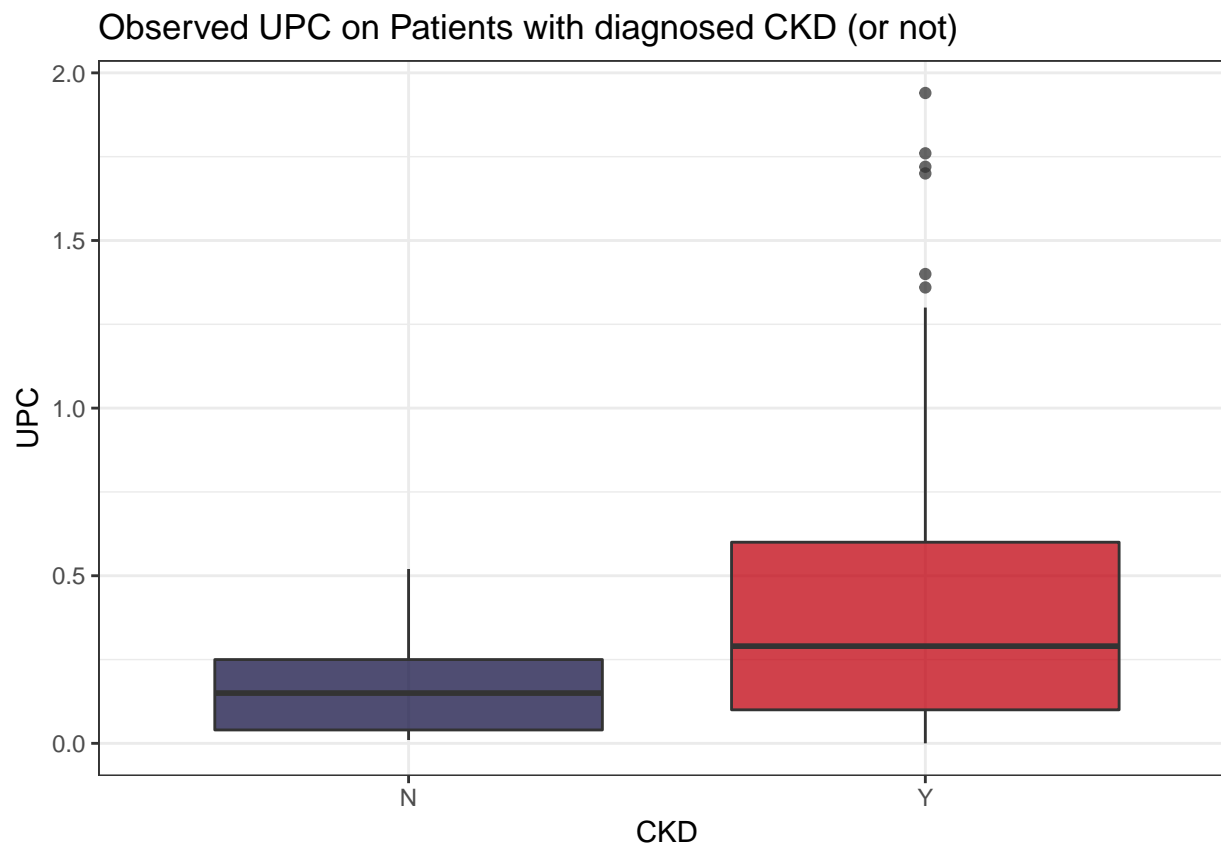
Sex



```
## # A tibble: 3 x 3
##   Sex   upc_mean upc_sd
##   <chr>   <dbl> <dbl>
## 1 F       0.731  1.79
## 2 H       0.526  0.802
## 3 M       0.497  0.624
```

```
## $msg1
## [1] "El resultat del test estadístic surt amb un p-value de 0.459. "
##
## $msg2
## [1] "Per tant la diferencia NO és estadísticament significativa."
```

CKD

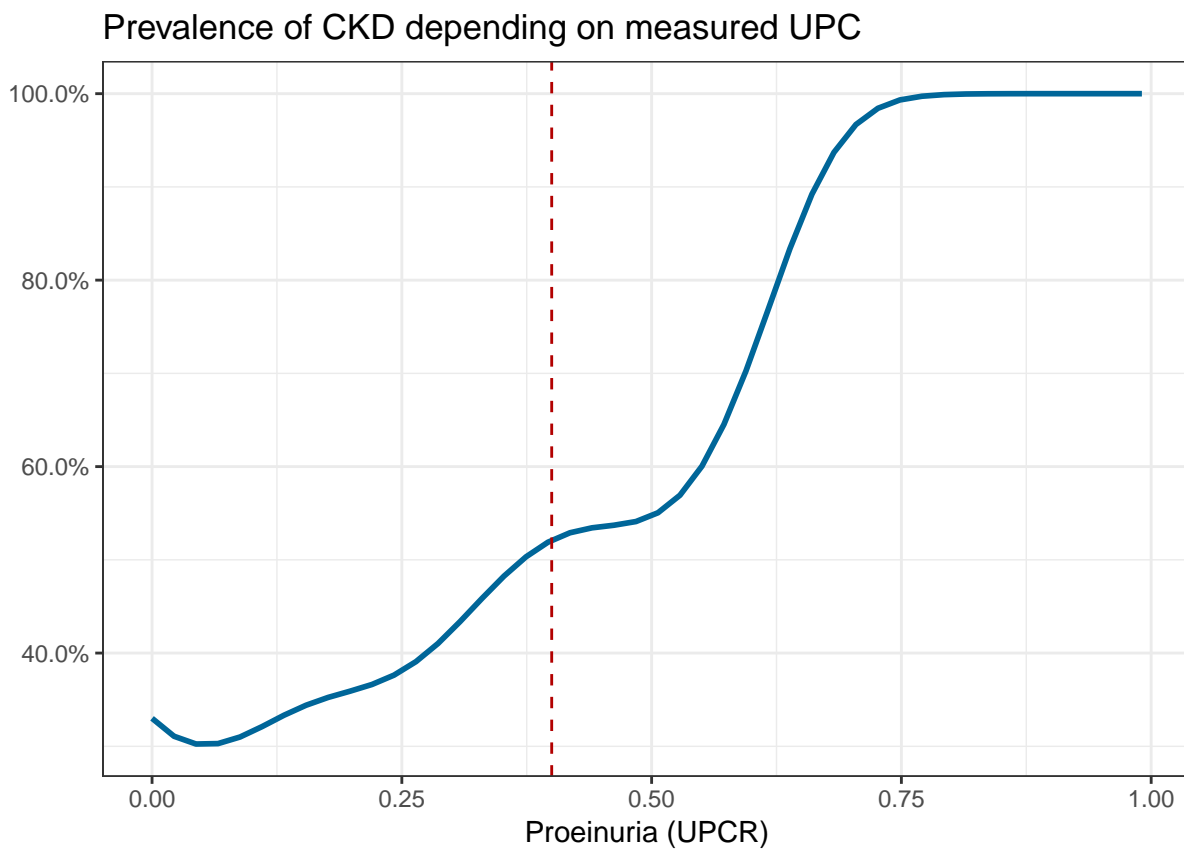


```
## # A tibble: 2 x 3
##   CKD   upc_mean upc_sd
##   <chr>   <dbl> <dbl>
## 1 N       0.257  0.444
## 2 Y       0.678  1.37
```

Aquí només mirem mitja i dispersió dels nivells de UPC mesurats, separant en pacients diagnosticats amb CKD i els que no. El que podem dir és que els pacients amb CKD diagnosticada tenen una UPC considerablement més alta (imagino que això seria prou obvi).

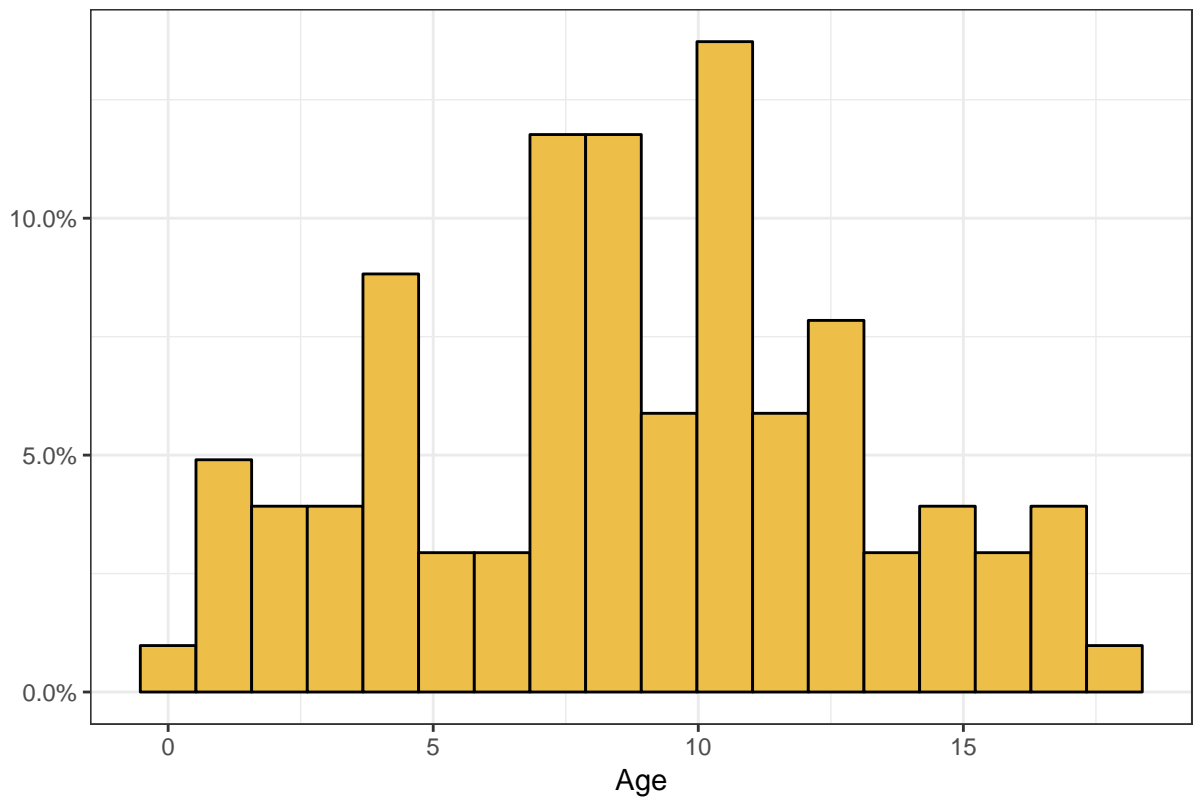
```
## [1] "El resultat del test estadístic surt amb un p-value de 0.027. "
```

```
## [1] "El que ens diu que la diferencia és estadísticament significativa."
```



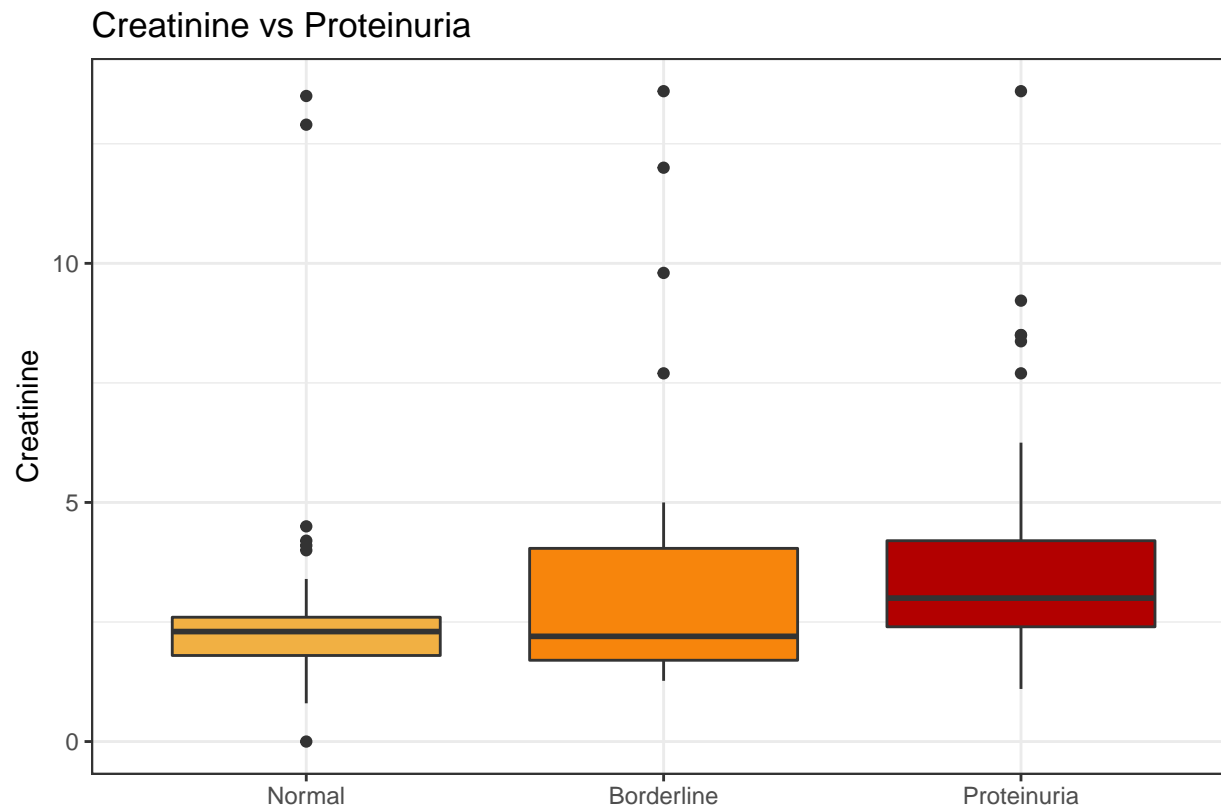
Més en detall he intentat explorar com es relaciona el nivell de UPC amb la diagnosi de CKD. El gràfic que veus ensenya el % de casos diagnosticats amb CKD en funció del UPC que presentaven. Crec que es prou rellevant que a partir de UPC superior a 0.5 es dispara cap al 100%. El que vol dir que amb proteinuria **Proteinuria** pràcticament tothom té CKD.

Age at CKD diagnosis



Encara imagino que això es pot estudiar més. Potser amb més dades. De moment lo que puc dir és que l'edat de diagnosi de CKD és més habitual entre els 7 i 11 anys. Però la diferencia tampoc és gaire rellevant

Creatinina, Urea & UPC

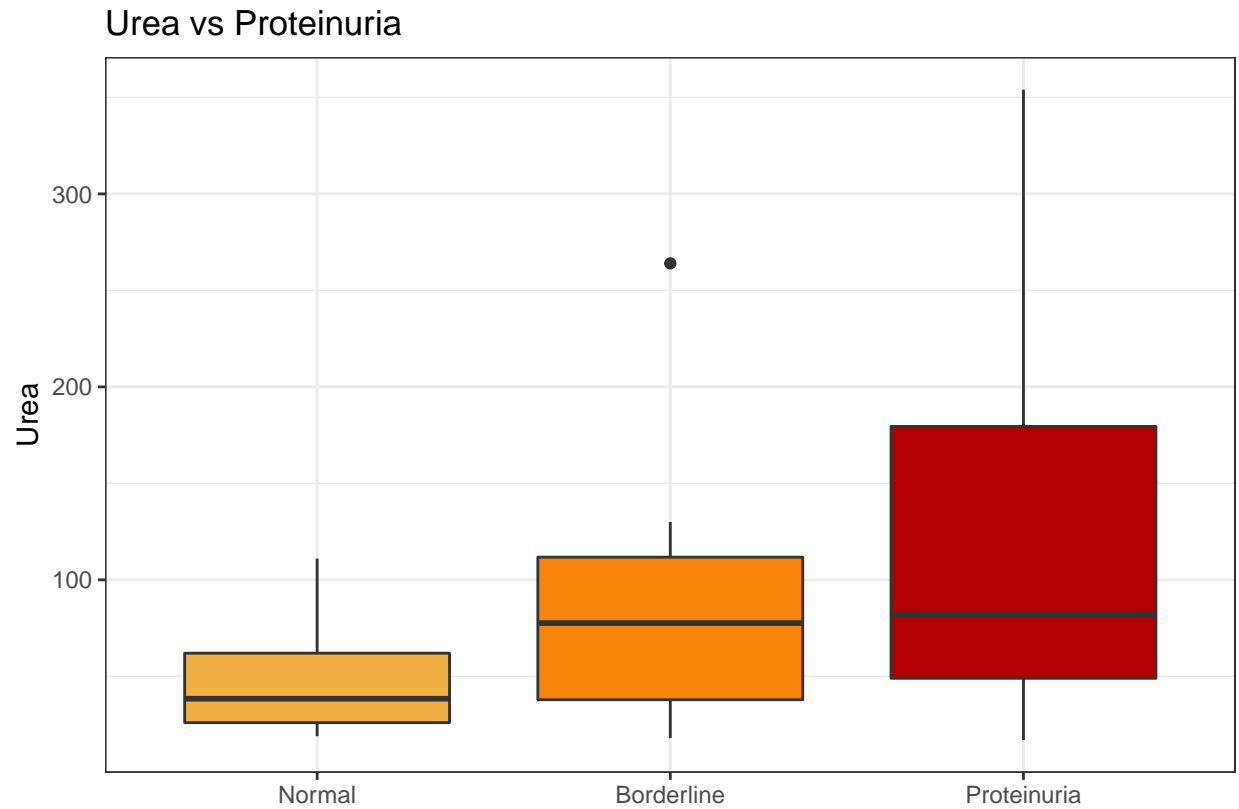


```
## # A tibble: 3 x 3
##   protein      crea_mean crea_sd
##   <fct>         <dbl>   <dbl>
## 1 Normal          2.79     2.52
## 2 Borderline       3.97     3.65
## 3 Proteinuria      3.91     2.74
```

Un altre boxplot. Veiem mitja i dispersió. Sembla prou evident que amb proteinuria **Proteinuria** s'observa una *Creatinina* més alta.

```
## [1] "El resultat del test estadístic surt amb un p-value de 0.077. "
```

```
## [1] "Per tant la diferencia NO és estadísticament significativa."
```



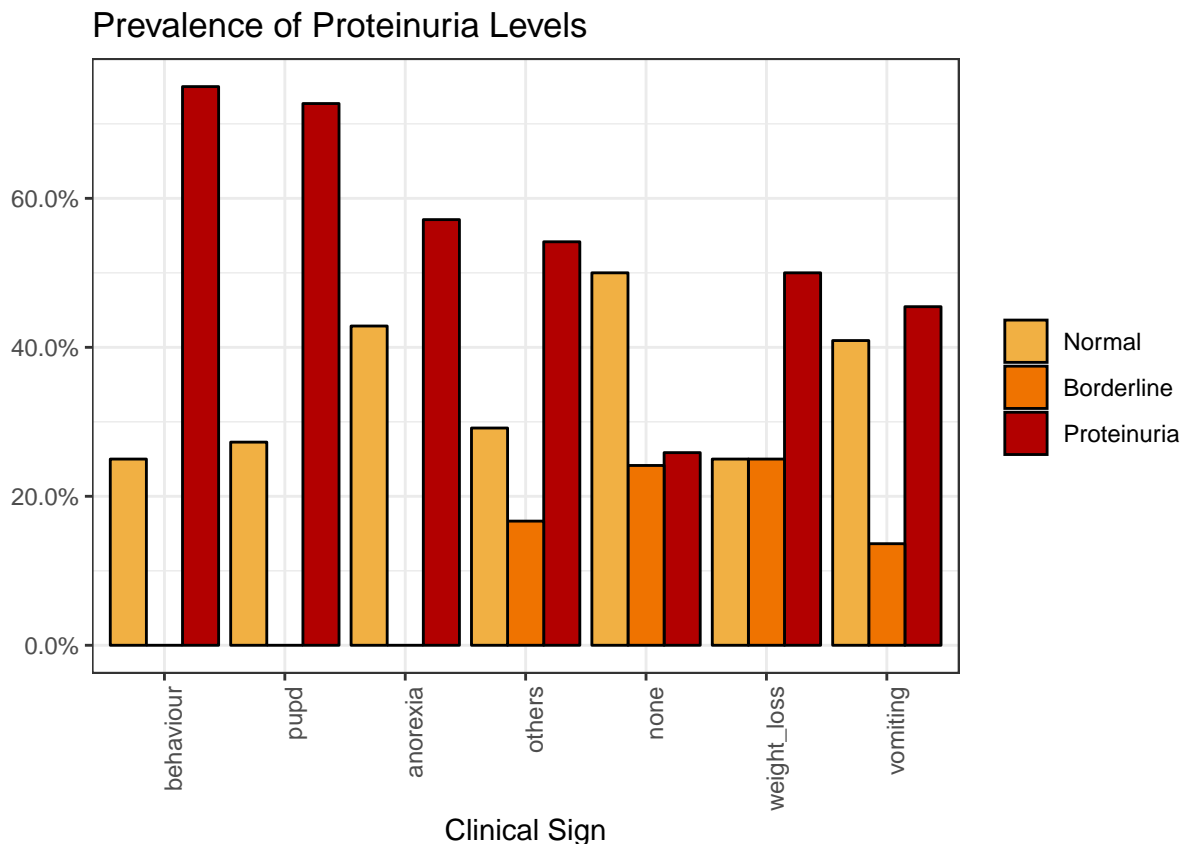
```
## # A tibble: 3 x 3
##   protein      urea_mean urea_sd
##   <fct>         <dbl>   <dbl>
## 1 Normal          45.0    23.3
## 2 Borderline      82.8    58.3
## 3 Proteinuria    115.    92.1
```

En el cas de la *Urea* observem més o menys el mateix, però en aquest cas l'impacte és molt més alt.

```
## [1] "El resultat del test estadístic surt amb un p-value de 0.002. "
```

```
## [1] "El que ens diu que la diferencia és estadísticament significativa."
```

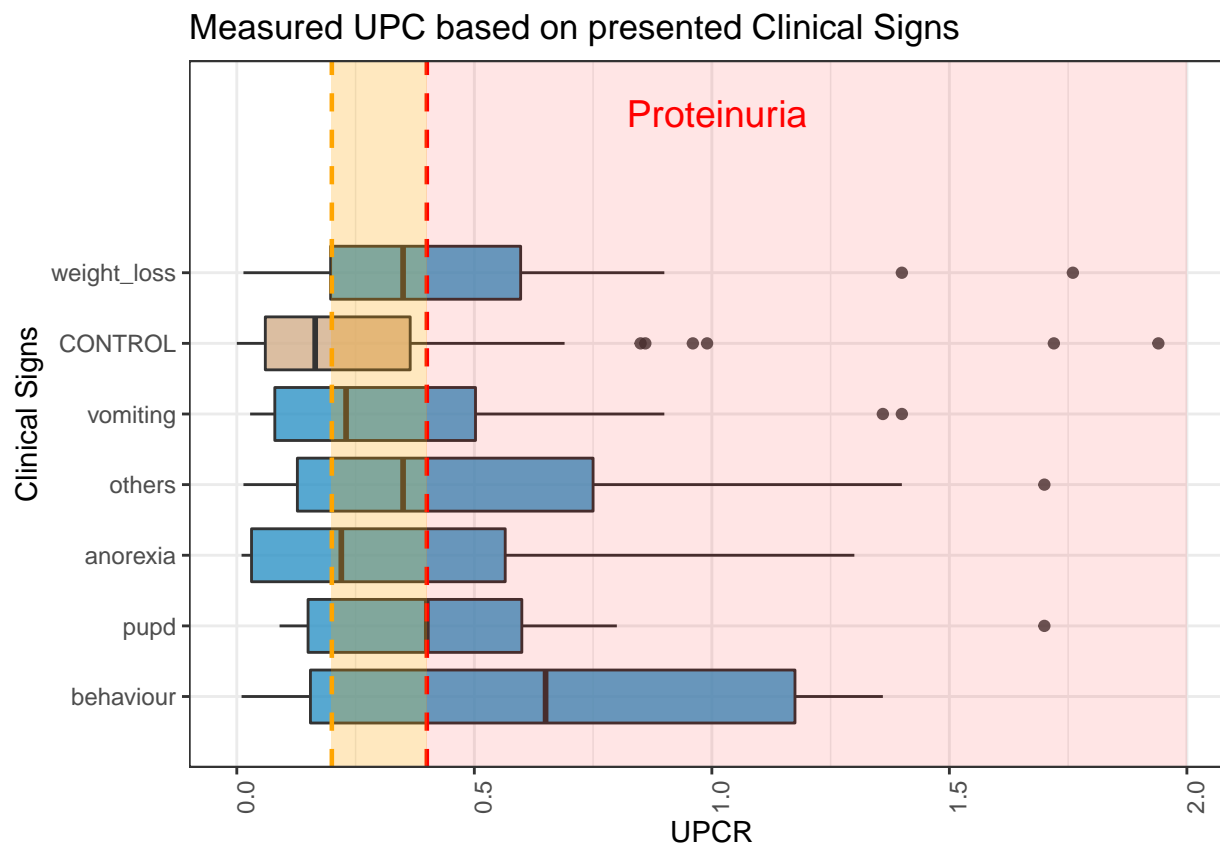
Clinical Signs



```
## # A tibble: 21 x 3
##   protein      variable    value
##   <fct>      <fct>      <dbl>
## 1 Proteinuria behaviour    0.75
## 2 Proteinuria pupd         0.727
## 3 Proteinuria anorexia     0.571
## 4 Proteinuria others       0.542
## 5 Normal      none         0.5
## 6 Proteinuria weight_loss  0.5
## 7 Proteinuria vomiting     0.455
## 8 Normal      anorexia     0.429
## 9 Normal      vomiting     0.409
## 10 Normal     others       0.292
## # ... with 11 more rows
```

Aquí m'he posat a investigar com podem relacionar els diferents *Clinical Signs* amb els nivells de proteinúria. Per tal que les dades tinguessin algun sentit. He reduït i agrupat moltes de les descripcions que es donaven en supergrups. Per exemple, he agrupat totes les descripcions com “acute vomiting” o “chronic vomiting” en la super-categoria “vomiting”. Sé que els detalls que es perden podrien dir coses interessants, però sense agrupar les descripcions tenim 100 *Clinical Signs* diferents per 100 pacients. Et passaré la llista de totes les agrupacions fetes, si hi ha alguna que voleu canviar, no em costa res ja. De moment ho he agrupat una mica com m'ha semblat.

FYI: El símptoma que he etiquetat com **behaviour** agrupa tot allò on a la descripció deia que l'animal es comportava estrany (lethargic, excited, weird behaviour, etc.)



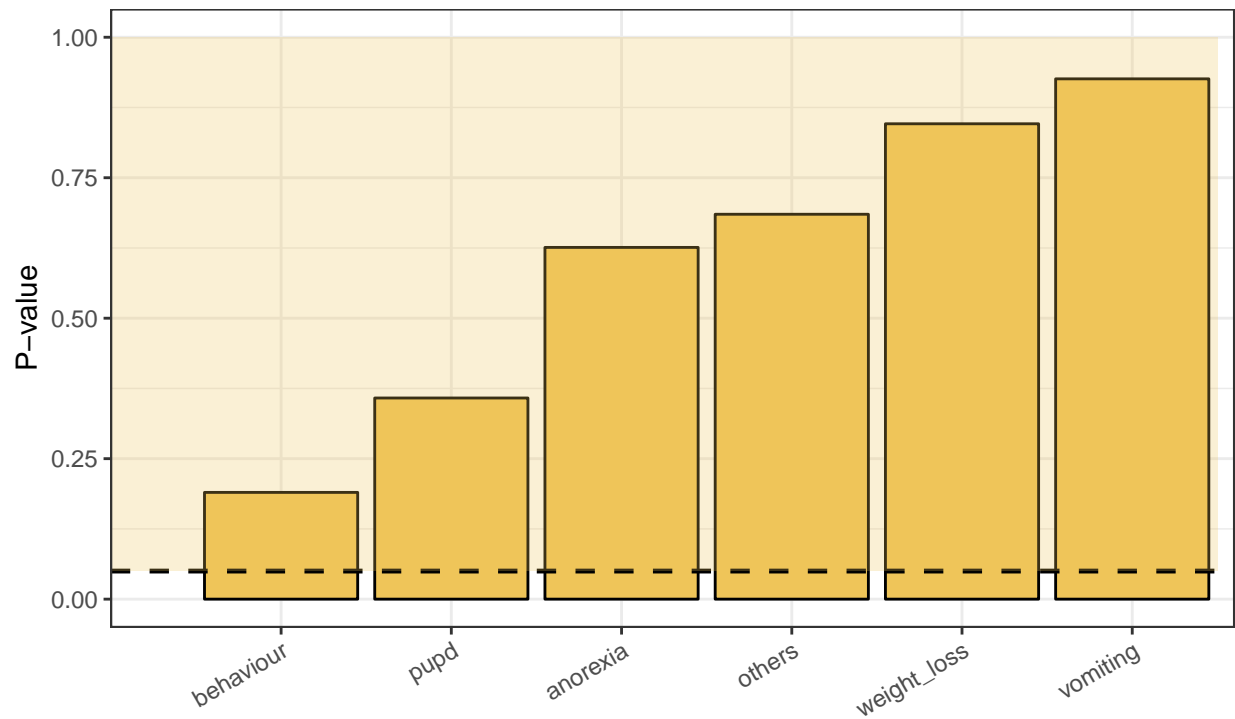
```
## # A tibble: 7 x 3
##   variable    upc_mean upc_sd
##   <fct>      <dbl>  <dbl>
## 1 anorexia    0.763  1.07
## 2 behaviour   1.06  0.873
## 3 none       0.554  1.54
## 4 pupd       0.86  0.860
## 5 vomiting   0.579  0.766
## 6 weight_loss 0.501  0.495
## 7 others     0.659  0.689
```

Semblant al gràfic anterior, però més detallat. Intento veure com els *Clinical Signs* que es presenten poden servir per “predir” el nivell de UPC que l'animal presentaria. He fet servir l'absència de cap símptoma (controls rutinaris o coses així) com a grup **control**. Digueu-me si això té sentit. Ho he posat per tenir una referència.

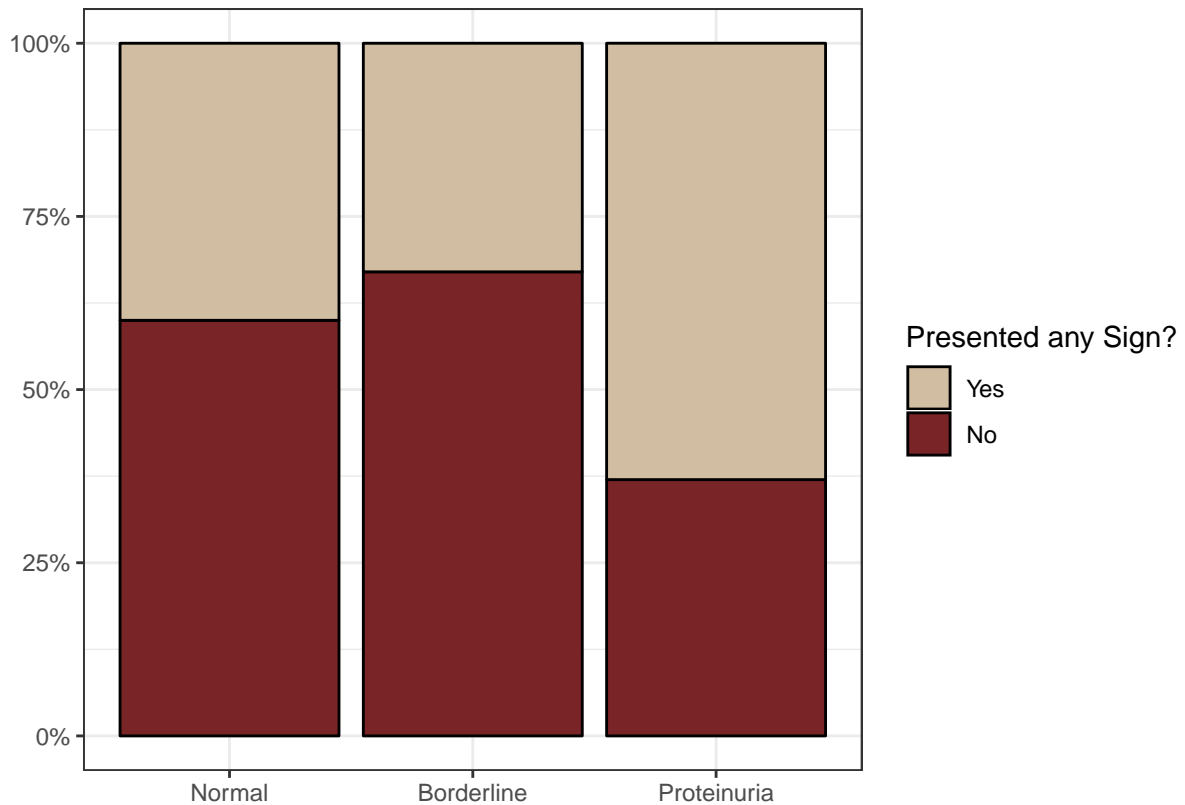
Aquests resultats els trobo interessants, per què cap pacient sense símptomes ha presentat nivells d'UPC que es classifiquen com proteinúria **Proteinuria**, tret d'uns pocs outliers.

```
## [1] 0.602
```

Results from Testing the difference between Clinical Signs and the Control Group



Cap test ha sigut significatiu. Tenim masses categories (tot i que les he reduït molt) pel vòlum de dades tan petit que tenim. Si fem el test diferenciant si té símptomes vs no en tenen cap, sí que tenim evidència significativa. Implica això que qualsevol símptoma ja és indicador de possible proteinúria? Jo abans diria que el resultat del test ens diu que no tenim prou dades com per disseminar entre més de dues categories.

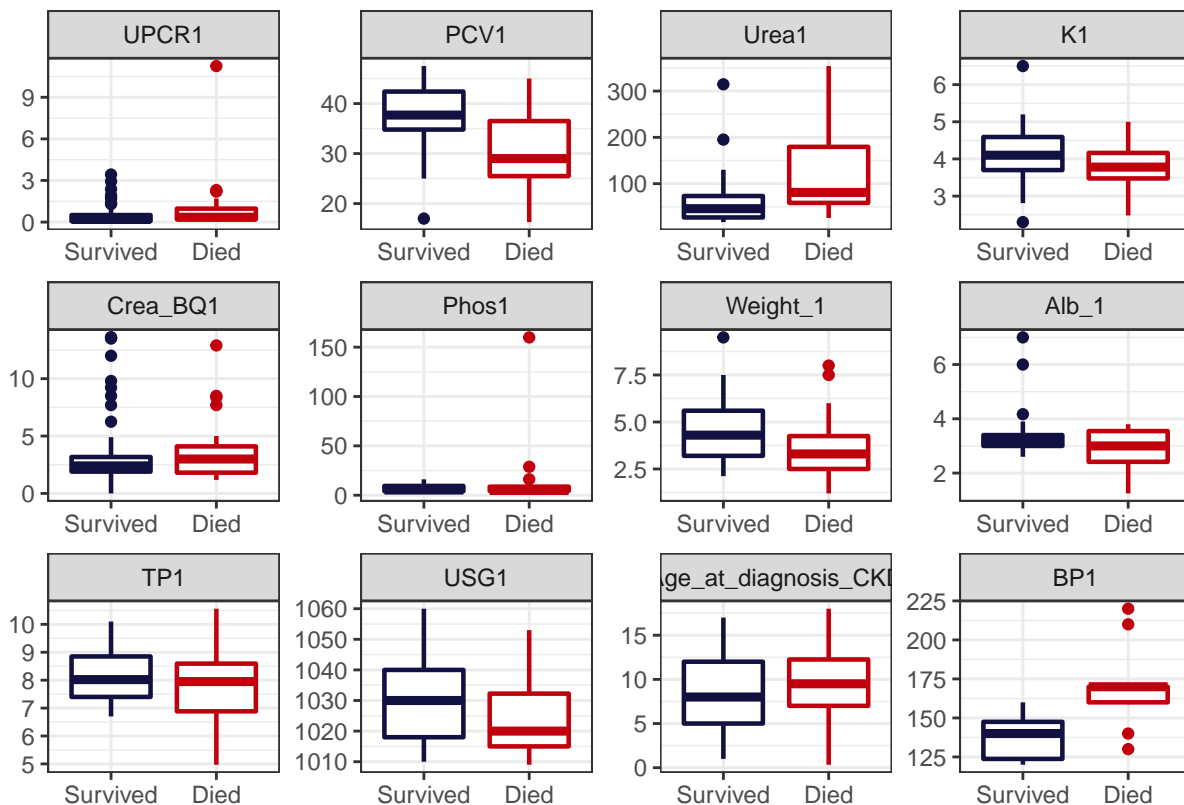


```
## # A tibble: 3 x 2
##   protein      n
##   <fct>      <dbl>
## 1 Normal      0.4
## 2 Borderline  0.33
## 3 Proteinuria 0.63
```

Aquest altre plot el que reflexa és tal qual quin percentatge de proeinurics presentàven cap signe clínic. A diferència de normals i borderlines, la majoria dels proteinúrics presentàven almenys algun signe clínic, el q fos.

Mortality

Ara em centrafe en analitzar quines de les variables que tenim poden tenir cert poder predictiu per la mortalitat.

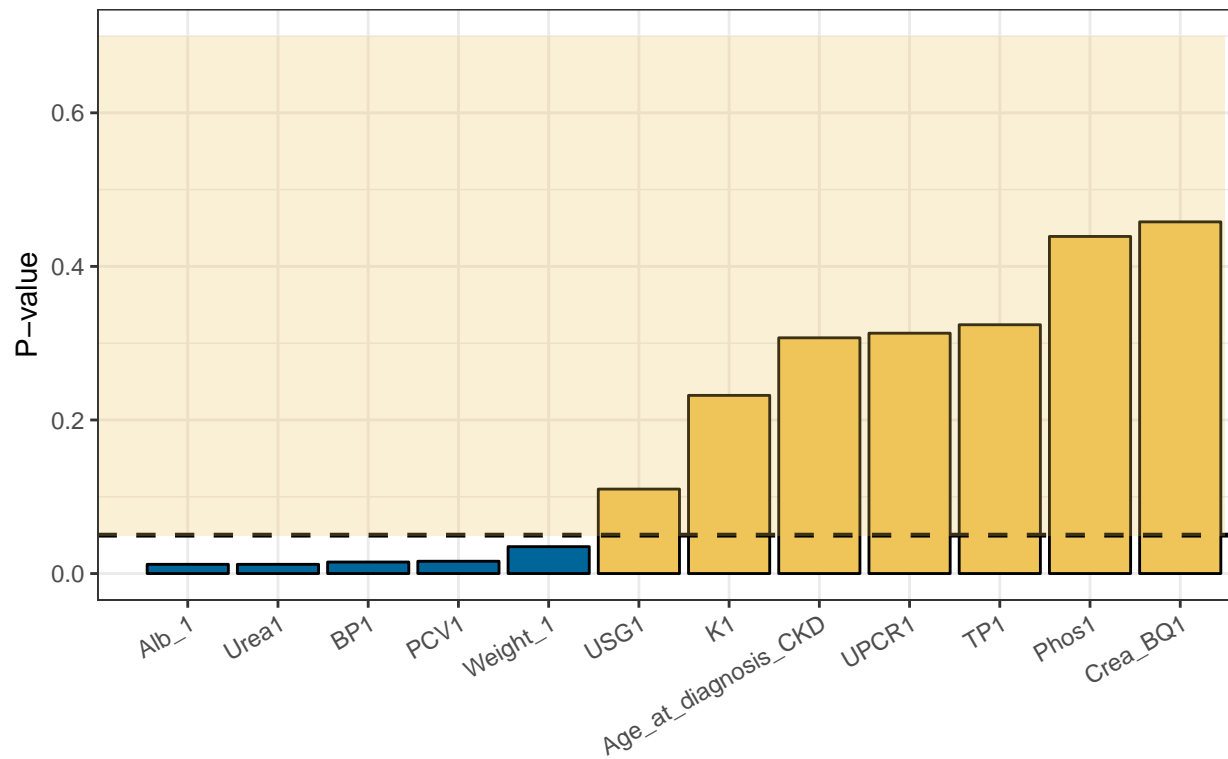


```
## # A tibble: 24 x 4
## # Groups:   death [?]
##   death variable value_mean value_sd
##   <lg1> <fct>      <dbl>    <dbl>
## 1 FALSE UPCR1      0.432    0.639
## 2 FALSE PCV1      37.2     7.90
## 3 FALSE Urea1     61.1    51.5
## 4 FALSE K1        4.11    0.907
## 5 FALSE Crea_BQ1  3.30    2.88
## 6 FALSE Phos1     7.00    3.99
## 7 FALSE Weight_1  4.56    1.66
## 8 FALSE Alb_1     3.34    0.730
## 9 FALSE TP1       8.15    0.904
## 10 FALSE USG1    1030    13.4
## # ... with 14 more rows
```

Aquí podem veure els boxplots de totes les variables que podem considerar per intentar predir la mortalitat dels pacients. Només mirant els gràfics ja podem predir que són rellevants la PCV, la Urea, la BP i el Potassi. De la resta o no és evident si són rellevants o no, o és evident que no ho són.

Per no haver-ho de fer a ull, podem fer un test estadístic a totes les comparacions.

Results from Hypothesis



```
## # A tibble: 24 x 5
## # Groups:   death [?]
##   death variable value_mean value_sd pvalues
##   <lg1> <fct>      <dbl>    <dbl> <dbl>
## 1 FALSE UPCR1      0.432    0.639 0.313
## 2 FALSE PCV1       37.2     7.90 0.016
## 3 FALSE Urea1      61.1    51.5 0.012
## 4 FALSE K1         4.11    0.907 0.232
## 5 FALSE Crea_BQ1    3.30    2.88 0.458
## 6 FALSE Phos1      7.00    3.99 0.439
## 7 FALSE Weight_1   4.56    1.66 0.035
## 8 FALSE Alb_1      3.34    0.730 0.012
## 9 FALSE TP1       8.15    0.904 0.324
## 10 FALSE USG1     1030    13.4 0.11
## # ... with 14 more rows
```

En aquest gràfic de barres podem veure el resultat del **pvalue** obtingut de tots els tests estadístics. El threshold s'ha situat a 0.05, que és l'estàndard presentat per Fisher. El que vol dir que totes aquelles variables amb un pvalue per sota d'aquest valor, podem dir que son estadísticament significatives, pel que respecta a la mortalitat.

Estadis CKD

