

# Homework 3

*Andey Nunes, Mengyu Li, Jordan Hilton, Peter Boss*

*2/2/2019*

## Document and Exercise Set Up

```
Claims_and_Days <- read_csv("Claims_and_Days.csv",
col_types = cols(CharlsonIndex = col_character()))
```

```
names(Claims_and_Days)
```

```
## [1] "ID" "MemberID"
## [3] "ProviderID" "vendor"
## [5] "pcp" "Year"
## [7] "specialty" "placesvc"
## [9] "paydelay" "LengthOfStay"
## [11] "dsfs" "PrimaryConditionGroup"
## [13] "CharlsonIndex" "sex"
## [15] "AgeAtFirstClaim" "DaysInHospital_Y2"
```

```
Claims_and_Days <- Claims_and_Days %>%
```

```
  mutate(CharlsonIndex = case_when(
    CharlsonIndex == "0" ~ "0",
    CharlsonIndex == "2-Jan" ~ "1-2",
    CharlsonIndex == "4-Mar" ~ "3-4",
    CharlsonIndex == "5+" ~ "5+"),
    AgeAtFirstClaim = case_when(
    AgeAtFirstClaim == "19-Oct" ~ "10-19",
    AgeAtFirstClaim != "19-Oct" ~ AgeAtFirstClaim),
    LengthOfStay = if_else(is.na(LengthOfStay), "0 or unknown", LengthOfStay)) %>%
  mutate(CharlsonIndex = ordered(CharlsonIndex, levels =
    c("0", "1-2", "3-4", "5+")),
    AgeAtFirstClaim = ordered(AgeAtFirstClaim, levels = c(
    "0-9", "10-19", "20-29", "30-39", "40-49", "50-59", "60-69", "70-79", "80+")),
    LengthOfStay = ordered(LengthOfStay, levels = c(
    "0 or unknown", "1 day", "2 days", "3 days", "4 days", "5 days", "6 days", "1- 2 weeks", "2- 4 weeks", "3- 6 weeks", "7- 12 weeks", "13- 24 weeks", "25- 50 weeks", "51- 100 weeks", "101- 200 weeks", "201- 400 weeks", "401- 800 weeks", "801- 1600 weeks", "1601- 3200 weeks", "3201- 6400 weeks", "6401- 12800 weeks", "12801- 25600 weeks", "25601- 51200 weeks", "51201- 102400 weeks", "102401- 204800 weeks", "204801- 409600 weeks", "409601- 819200 weeks", "819201- 1638400 weeks", "1638401- 3276800 weeks", "3276801- 6553600 weeks", "6553601- 13107200 weeks", "13107201- 26214400 weeks", "26214401- 52428800 weeks", "52428801- 104857600 weeks", "104857601- 209715200 weeks", "209715201- 419430400 weeks", "419430401- 838860800 weeks", "838860801- 1677721600 weeks", "1677721601- 3355443200 weeks", "3355443201- 6710886400 weeks", "6710886401- 13421772800 weeks", "13421772801- 26843545600 weeks", "26843545601- 53687091200 weeks", "53687091201- 107374182400 weeks", "107374182401- 214748364800 weeks", "214748364801- 429496729600 weeks", "429496729601- 858993459200 weeks", "858993459201- 1717986918400 weeks", "1717986918401- 3435973836800 weeks", "3435973836801- 6871947673600 weeks", "6871947673601- 13743895347200 weeks", "13743895347201- 27487790694400 weeks", "27487790694401- 54975581388800 weeks", "54975581388801- 109951162777600 weeks", "109951162777601- 219902325555200 weeks", "219902325555201- 439804651110400 weeks", "439804651110401- 879609302220800 weeks", "879609302220801- 1759218604441600 weeks", "1759218604441601- 3518437208883200 weeks", "3518437208883201- 7036874417766400 weeks", "7036874417766401- 14073748835532800 weeks", "14073748835532801- 28147497671065600 weeks", "28147497671065601- 56294995342131200 weeks", "56294995342131201- 112589990684262400 weeks", "112589990684262401- 225179981368524800 weeks", "225179981368524801- 450359962737049600 weeks", "450359962737049601- 900719925474099200 weeks", "900719925474099201- 1801439850948198400 weeks", "1801439850948198401- 3602879701896396800 weeks", "3602879701896396801- 7205759403792793600 weeks", "7205759403792793601- 14411518807585587200 weeks", "14411518807585587201- 28823037615171174400 weeks", "28823037615171174401- 57646075230342348800 weeks", "57646075230342348801- 115292150460684697600 weeks", "115292150460684697601- 230584300921369395200 weeks", "230584300921369395201- 461168601842738790400 weeks", "461168601842738790401- 922337203685477580800 weeks", "922337203685477580801- 1844674407370955161600 weeks", "1844674407370955161601- 3689348814741910323200 weeks", "3689348814741910323201- 7378697629483820646400 weeks", "7378697629483820646401- 14757395258967641292800 weeks", "14757395258967641292801- 29514790517935282585600 weeks", "29514790517935282585601- 59029581035870565171200 weeks", "59029581035870565171201- 118059162071741130342400 weeks", "118059162071741130342401- 236118324143482260684800 weeks", "236118324143482260684801- 472236648286964521369600 weeks", "472236648286964521369601- 944473296573929042739200 weeks", "944473296573929042739201- 1888946593147858085478400 weeks", "1888946593147858085478401- 3777893186295716170956800 weeks", "3777893186295716170956801- 7555786372591432341913600 weeks", "7555786372591432341913601- 15111572745182864683827200 weeks", "15111572745182864683827201- 30223145490365729367654400 weeks", "30223145490365729367654401- 60446290980731458735308800 weeks", "60446290980731458735308801- 120892581961462917470617600 weeks", "120892581961462917470617601- 241785163922925834941235200 weeks", "241785163922925834941235201- 483570327845851669882470400 weeks", "483570327845851669882470401- 967140655691703339764940800 weeks", "967140655691703339764940801- 1934281311383406679529881600 weeks", "1934281311383406679529881601- 3868562622766813359059763200 weeks", "3868562622766813359059763201- 7737125245533626718119526400 weeks", "7737125245533626718119526401- 15474250491067253436239052800 weeks", "15474250491067253436239052801- 30948500982134506872478105600 weeks", "30948500982134506872478105601- 61897001964269013744956211200 weeks", "61897001964269013744956211201- 123794003928538027489912422400 weeks", "123794003928538027489912422401- 247588007857076054979824844800 weeks", "247588007857076054979824844801- 495176015714152109959649689600 weeks", "495176015714152109959649689601- 990352031428304219919299379200 weeks", "990352031428304219919299379201- 1980704062856608439838598758400 weeks", "1980704062856608439838598758401- 3961408125713216879677197516800 weeks", "3961408125713216879677197516801- 7922816251426433759354395033600 weeks", "7922816251426433759354395033601- 15845632502852867518708790067200 weeks", "15845632502852867518708790067201- 31691265005705735037417580134400 weeks", "31691265005705735037417580134401- 63382530011411470074835160268800 weeks", "63382530011411470074835160268801- 126765060022822940149670320537600 weeks", "126765060022822940149670320537601- 253530120045645880299340641075200 weeks", "253530120045645880299340641075201- 507060240091291760598681282150400 weeks", "507060240091291760598681282150401- 1014120480182583521197362564300800 weeks", "1014120480182583521197362564300801- 2028240960365167042394725128601600 weeks", "2028240960365167042394725128601601- 4056481920730334084789450257203200 weeks", "4056481920730334084789450257203201- 8112963841460668169578900514406400 weeks", "8112963841460668169578900514406401- 16225927682921336339157801028812800 weeks", "16225927682921336339157801028812801- 32451855365842672678315602057625600 weeks", "32451855365842672678315602057625601- 64903710731685345356631204115251200 weeks", "64903710731685345356631204115251201- 129807421463370690713262408230502400 weeks", "129807421463370690713262408230502401- 259614842926741381426524816461004800 weeks", "259614842926741381426524816461004801- 519229685853482762853049632922009600 weeks", "519229685853482762853049632922009601- 1038459371706965525706099265844019200 weeks", "1038459371706965525706099265844019201- 2076918743413931051412198531688038400 weeks", "2076918743413931051412198531688038401- 4153837486827862102824397063376076800 weeks", "4153837486827862102824397063376076801- 8307674973655724205648794126752153600 weeks", "8307674973655724205648794126752153601- 16615349947311448411297588253504307200 weeks", "16615349947311448411297588253504307201- 33230699894622896822595176507008614400 weeks", "33230699894622896822595176507008614401- 66461399789245793645190353014017228800 weeks", "66461399789245793645190353014017228801- 132922799578491587290380706028034457600 weeks", "132922799578491587290380706028034457601- 265845599156983174580761412056068915200 weeks", "265845599156983174580761412056068915201- 531691198313966349161522824112137830400 weeks", "531691198313966349161522824112137830401- 1063382396627932698323045648224275660800 weeks", "1063382396627932698323045648224275660801- 2126764793255865396646091296448551321600 weeks", "2126764793255865396646091296448551321601- 4253529586511730793292182592897102643200 weeks", "4253529586511730793292182592897102643201- 8507059173023461586584365185794205286400 weeks", "8507059173023461586584365185794205286401- 17014118346046923173168730371588410572800 weeks", "17014118346046923173168730371588410572801- 34028236692093846346337460743176821145600 weeks", "34028236692093846346337460743176821145601- 68056473384187692692674921486353642291200 weeks", "68056473384187692692674921486353642291201- 136112946768375385385349842972707284582400 weeks", "136112946768375385385349842972707284582401- 272225893536750770770699685945414569164800 weeks", "272225893536750770770699685945414569164801- 544451787073501541541399371890829138329600 weeks", "544451787073501541541399371890829138329601- 1088903574147003083082798743781658276659200 weeks", "1088903574147003083082798743781658276659201- 2177807148294006166165597487563316553318400 weeks", "2177807148294006166165597487563316553318401- 4355614296588012332331194975126633106636800 weeks", "4355614296588012332331194975126633106636801- 8711228593176024664662389950253266213273600 weeks", "8711228593176024664662389950253266213273601- 17422457186352049329324779900506532426547200 weeks", "17422457186352049329324779900506532426547201- 34844914372704098658649559801013064853094400 weeks", "34844914372704098658649559801013064853094401- 69689828745408197317299119602026129706188800 weeks", "69689828745408197317299119602026129706188801- 139379657490816394634598239204052259412377600 weeks", "139379657490816394634598239204052259412377601- 278759314981632789269196478408104518824755200 weeks", "278759314981632789269196478408104518824755201- 557518629963265578538392956816209037649510400 weeks", "557518629963265578538392956816209037649510401- 1115037259926531157076785913632418075299020800 weeks", "1115037259926531157076785913632418075299020801- 2230074519853062314153571827264836150598041600 weeks", "2230074519853062314153571827264836150598041601- 4460149039706124628307143654529672301196083200 weeks", "4460149039706124628307143654529672301196083201- 8920298079412249256614287309059344602392166400 weeks", "8920298079412249256614287309059344602392166401- 17840596158824498513228574618118689204784332800 weeks", "17840596158824498513228574618118689204784332801- 35681192317648997026457149236237378409568665600 weeks", "35681192317648997026457149236237378409568665601- 71362384635297994052914298472474756819137331200 weeks", "71362384635297994052914298472474756819137331201- 142724769270595988105828596944949513638274662400 weeks", "142724769270595988105828596944949513638274662401- 285449538541191976211657193889899027276549324800 weeks", "285449538541191976211657193889899027276549324801- 570899077082383952423314387779798054553098649600 weeks", "570899077082383952423314387779798054553098649601- 1141798154164767904846628775559596109106197299200 weeks", "1141798154164767904846628775559596109106197299201- 2283596308329535809693257551119192218212394598400 weeks", "2283596308329535809693257551119192218212394598401- 4567192616659071619386515102238384436424789196800 weeks", "4567192616659071619386515102238384436424789196801- 9134385233318143238773030204476768872849578393600 weeks", "9134385233318143238773030204476768872849578393601- 18268770466636286477546060408953537745699156787200 weeks", "18268770466636286477546060408953537745699156787201- 36537540933272572955092120817907075491398313574400 weeks", "36537540933272572955092120817907075491398313574401- 73075081866545145910184241635814150982796627148800 weeks", "73075081866545145910184241635814150982796627148801- 146150163733090291820368483271628301965593254297600 weeks", "146150163733090291820368483271628301965593254297601- 292300327466180583640736966543256603931186508595200 weeks", "292300327466180583640736966543256603931186508595201- 584600654932361167281473933086513207862373017190400 weeks", "584600654932361167281473933086513207862373017190401- 1169201309864722334562947866173026415724746034380800 weeks", "1169201309864722334562947866173026415724746034380801- 2338402619729444669125895732346052831449492068761600 weeks", "2338402619729444669125895732346052831449492068761601- 4676805239458889338251791464692105662898984137523200 weeks", "4676805239458889338251791464692105662898984137523201- 9353610478917778676503582929384211325797968275046400 weeks", "9353610478917778676503582929384211325797968275046401- 18707220957835557353007165858768422651595936550092800 weeks", "18707220957835557353007165858768422651595936550092801- 37414441915671114706014331717536845303191873100185600 weeks", "37414441915671114706014331717536845303191873100185601- 74828883831342229412028663435073690606383746200371200 weeks", "74828883831342229412028663435073690606383746200371201- 149657767662684458824057326870147381212767492400742400 weeks", "149657767662684458824057326870147381212767492400742401- 299315535325368917648114653740294762425534984801484800 weeks", "299315535325368917648114653740294762425534984801484801- 598631070650737835296229307480589524851069969602969600 weeks", "598631070650737835296229307480589524851069969602969601- 1197262141301475670592458614961179049702139939205939200 weeks", "1197262141301475670592458614961179049702139939205939201- 2394524282602951341184917229922358099404279878411878400 weeks", "2394524282602951341184917229922358099404279878411878401- 4789048565205902682369834459844716198808559756823756800 weeks", "4789048565205902682369834459844716198808559756823756801- 9578097130411805364739668919689432397617119513647513600 weeks", "9578097130411805364739668919689432397617119513647513601- 19156194260823610729479337839378864795234239027295027200 weeks", "19156194260823610729479337839378864795234239027295027201- 38312388521647221458958675678757729590468478054590054400 weeks", "38312388521647221458958675678757729590468478054590054401- 76624777043294442917917351357515459180936956109180108800 weeks", "76624777043294442917917351357515459180936956109180108801- 153249554086588885835834702
```

```

    Risk_Level == "A" ~ "no risk",
    Risk_Level == "B" ~ "very low risk",
    Risk_Level == "C" ~ "low risk",
    Risk_Level == "D" ~ "medium risk",
    Risk_Level == "E" ~ "high risk")) %>%
mutate(Risk_Level_label = ordered(Risk_Level_label, levels = c(
  "no risk", "very low risk", "low risk", "medium risk", "high risk")))

str(VlookupSim)

## Classes 'tbl_df', 'tbl' and 'data.frame':   644706 obs. of  18 variables:
## $ ID : num  2 3 4 5 6 7 1 8 9 10 ...
## $ MemberID : num  25872 25872 25872 25872 25872 ...
## $ ProviderID : num  376108719 171278567 171278567 171278567 171278567 ...
## $ vendor : num  5024957 7891165 7891165 7891165 7891165 ...
## $ pcip : num  294037 294037 294037 294037 294037 ...
## $ Year : chr  "Y1" "Y1" "Y1" "Y1" ...
## $ specialty : chr  "Laboratory" "Internal" "Internal" "Internal" ...
## $ placesvc : chr  "Independent Lab" "Office" "Office" "Office" ...
## $ paydelay : num  23 16 19 21 21 11 22 37 23 30 ...
## $ LengthOfStay : Ord.factor w/ 13 levels "0 or unknown"<...: 1 1 1 1 1 1 1 1 1 1 ...
## $ dsfs : chr  "0- 1 month" "1- 2 months" "2- 3 months" "3- 4 months" ...
## $ PrimaryConditionGroup: chr  "MSC2a3" "RESPR4" "RESPR4" "RESPR4" ...
## $ CharlsonIndex : Ord.factor w/ 4 levels "0"<"1-2"<"3-4"<...: 1 2 2 2 2 2 2 1 1 1 ...
## $ sex : chr  "F" "F" "F" "F" ...
## $ AgeAtFirstClaim : Ord.factor w/ 9 levels "0-9"<"10-19"<...: 2 2 2 2 2 2 2 9 9 9 ...
## $ DaysInHospital_Y2 : num  0 0 0 0 0 0 0 0 0 0 ...
## $ Risk_Level : chr  "A" "A" "A" "A" ...
## $ Risk_Level_label : Ord.factor w/ 5 levels "no risk"<"very low risk"<...: 1 1 1 1 1 1 1 1 1 1 .

```

## Create pivot tables

### proportion of observations for risk groups

```

prop_obs <- function(df = VlookupSim, x,
                     y = VlookupSim$Risk_Level_label) {
  prop <- with(df, table(x, y))
  prop <- cbind(prop, total = rowSums(prop))
  high_prop <- prop[,5]/prop[,6]*100
  prop <- cbind(prop, high_risk_proportion = high_prop)
  #proptibble <- tibble(prop) %>%
  # arrange(desc(high_risk_proportion))

  return(prop)
}

```

### Primary Condition Group Counts and Proportion of High Risk

```

pcg <- prop_obs(x = VlookupSim$PrimaryConditionGroup)
kable(pcg)

```

	no risk	very low risk	low risk	medium risk	high risk	total	high_risk_proportion
AMI	5919	713	483	588	1188	8891	13.4

	no risk	very low risk	low risk	medium risk	high risk	total	high_risk_proportion
APPCHOL	3805	461	207	213	355	5041	7.0
ARTHSPIN	54575	6096	3087	3080	4873	71711	6.8
CANCRA	866	80	19	35	101	1101	9.2
CANCRB	7343	685	346	421	944	9739	9.7
CANCRM	170	21	11	7	22	231	9.5
CATAST	309	33	16	19	68	445	15.3
CHF	2004	322	130	153	563	3172	17.8
COPD	8516	1007	409	575	1199	11706	10.2
FLaELEC	899	104	45	55	144	1247	11.6
FXDISLC	6951	710	289	262	470	8682	5.4
GIBLEED	21002	2755	1343	1222	2524	28846	8.8
GIOBSENT	2186	230	132	94	189	2831	6.7
GYNEC1	8793	775	734	553	379	11234	3.4
GYNECA	1847	162	163	91	106	2369	4.5
HEART2	8446	1124	542	716	1591	12419	12.8
HEART4	5142	612	327	270	726	7077	10.3
HEMTOL	4513	461	177	286	757	6194	12.2
HIPFX	655	106	82	50	130	1023	12.7
INFEC4	17064	1864	783	691	1139	21541	5.3
LIVERDZ	595	57	20	30	48	750	6.4
METAB1	752	95	35	56	86	1024	8.4
METAB3	56338	5377	2383	2662	5040	71800	7.0
MISCHRT	24141	2585	994	1264	2437	31421	7.8
MISCL1	968	125	64	46	98	1301	7.5
MISCL5	9330	1098	496	524	867	12315	7.0
MSC2a3	89147	8331	3899	3769	6353	111499	5.7
NEUMENT	35789	3936	1524	1713	3343	46305	7.2
ODaBNCA	10050	844	368	397	657	12316	5.3
PERINTL	128	9	8	6	10	161	6.2
PERVALV	618	61	32	54	85	850	10.0
PNCRDZ	179	40	0	20	20	259	7.7
PNEUM	1905	238	81	131	226	2581	8.8
PRGNCY	5336	411	1048	609	429	7833	5.5
RENAL1	58	35	9	4	18	124	14.5
RENAL2	1323	178	76	97	429	2103	20.4
RENAL3	9657	1070	496	554	972	12749	7.6
RESPR4	29576	3274	1355	1266	2062	37533	5.5
ROAMI	8884	1200	622	676	1453	12835	11.3
SEIZURE	3530	625	271	278	458	5162	8.9
SEPSIS	76	12	5	10	17	120	14.2
SKNAUT	21217	2150	847	903	1784	26901	6.6
STROKE	1474	207	122	110	236	2149	11.0
TRAUMA	14868	1752	710	607	1023	18960	5.4
UTI	7817	822	403	433	680	10155	6.7

### Charlson Index Group Counts and Proportion of High Risk

```
ci <- prop_obs(x = VlookupSim$CharlsonIndex)
kable(ci)
```

	no risk	very low risk	low risk	medium risk	high risk	total	high_risk_proportion
0	299915	27731	13408	12086	16051	369191	4.3
1-2	186875	24094	11320	12828	28407	263524	10.8
3-4	7136	902	414	627	1701	10780	15.8
5+	835	126	51	59	140	1211	11.6

### Length of Stay Counts and Proportion of High Risk

```
los <- prop_obs(x = VlookupSim$LengthOfStay)
kable(los)
```

### Simulate the 1R Algorithm

```
data1R <- filter(VlookupSim, Risk_Level != "A")
```

Ignoring Risk Level A build the 1R Rule for each of

- “PrimaryConditionGroup”
- “CharlsonIndex”
- “LengthOfStay”

Select the Best Rule and Paste in here

### Questions:

#### Q4.1

Why does our selected rule work better (reference error rate)

#### Q4.2

For the 1R Rule is it better to have a lower or higher cardinality?

#### Q4.3

We are ignoring risk level A because it is a subset of the data that is independent of the group we are trying to predict and it consists of nearly 77% of the original observations. By working with a smaller subset that contains only the observations with any risk level other than “no risk”, we have a smaller search space for our algorithm.

### Extra Credit

Ignoring Risk Level A rebin the risk level buckets and rebuild the 1R Rule and see if there is a difference.

# Appendix

## Risk Group Counts

graphical inspection function

```
risk_group_counts <- ggplot(VlookupSim, aes(x = Risk_Level_label)) +  
  geom_text(aes(label = ..count..), stat = "count", vjust = -0.25) +  
  geom_bar()
```

```
risk_group_counts +  
  facet_wrap(~PrimaryConditionGroup)
```

```
risk_group_counts +  
  facet_wrap(~CharlsonIndex)
```

```
risk_group_counts +  
  facet_wrap(~LengthOfStay)
```