Capstone EDA

Gabby Lopez

2025-01-28

There are 673 patients that are marked as having prediabetes as well as type II diabetes. Not a big deal for the actual analysis itself since we will combine these two indicator variables into one variable, but we should clear this up for our descriptive statistics.

table 1 for KOH only - need to add average number of KOH meetings attended

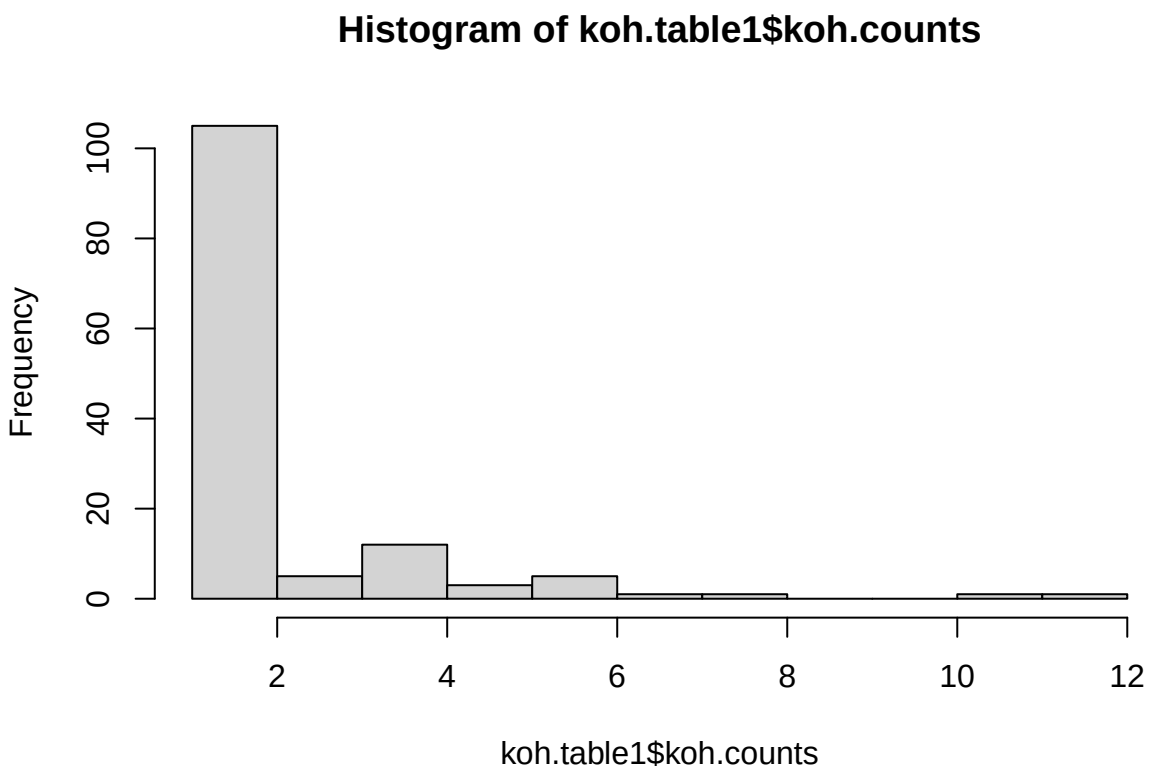
```
## Adding missing grouping variables: `UniqueIdentifier`
```

Baseline Characteristics	**KOH Participants** N = 136
Age	
Median (Min, Max)	58 (21, 80)
Sex	
Female	89 (65%)
Male	47 (35%)
KOH Meetings Attended	
Median (Min, Max)	1 (1, 12)
Missing	2
Hypertension	49 (36%)
Pre-Diabetes or T2DM	97 (71%)
Hypertension and T2DM	47 (35%)
Clinic Location	
Maple	24 (28%)
Market	62 (72%)
Missing	50
Housing Status	
Housed	70 (51%)
Doubling Up	64 (47%)
Permanent Supportive Housing	1 (0.7%)
Transitional	0 (0%)
Homeless Shelter	1 (0.7%)

Baseline Characteristics	KOH Participants N = 136
Street	0 (0%)
Other	0 (0%)
Unknown	0 (0%)
Income Level	
Median (Min, Max)	5 (0, 294)
Missing	8
Risk Score	
Median (Min, Max)	1 (0, 9)
Missing	6

distribution of KOH attendance

```
hist(koh.table1$koh.counts)
```




```
summary(koh.table1$koh.counts)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.000   1.000   2.045   2.000  12.000       2
```

```
summary(as.factor(koh.table1$koh.counts))
```

```
##    1    2    3    4    5    6    7    8   11   12 NA's
##   82   23    5   12    3    5    1    1    1    1    2
```

subsetting koh patients with hypertension

```
# get pts with htn
koh.htn <- koh.table1 %>% filter(HTN == 1)
# isolate the patient ids
koh.htn.patlist <- koh.htn %>% select("UniqueIdentifier")
# select bp measures for only kohn htn pts
koh.bp <- left_join(koh.htn.patlist, bp.nona.18, by = "UniqueIdentifier") %>% select(-age)

# count number of bp readings per patient
koh.bp.counts <- koh.bp %>% group_by(UniqueIdentifier) %>% count(name = "bp.counts")
# remove any patients that only have 1 reading - i will filter out any that don't have reading before a
koh.htn.elig.patlist <- koh.bp.counts %>% filter(bp.counts > 1) %>% select("UniqueIdentifier")
# now that we have the "eligible patients" we can subset the BP data for these patients
koh.bp.elig <- left_join(koh.htn.elig.patlist, koh.bp, by = "UniqueIdentifier")
```

subsetting koh patients with prediabetes or T2DM

```
# get pts with diabetes
koh.t2dm <- koh.table1 %>% filter(Diabetes == 1)
# isolate the patient ids
koh.t2dm.patlist <- koh.t2dm %>% select("UniqueIdentifier")
# select a1c measures for only koh t2dm pts
koh.a1c <- left_join(koh.t2dm.patlist, a1c.nona.18, by = "UniqueIdentifier") %>% select(-age)

# count number of a1c readings per patient
koh.a1c.counts <- koh.a1c %>% group_by(UniqueIdentifier) %>% count(name = "a1c.counts")
# remove any patients that only have 1 reading - i will filter out any that don't have reading before a
koh.t2dm.elig.patlist <- koh.a1c.counts %>% filter(a1c.counts > 1) %>% select("UniqueIdentifier")
# now that we have the "eligible patients" we can subset the a1c data for these patients
koh.a1c.elig <- left_join(koh.t2dm.elig.patlist, koh.a1c, by = "UniqueIdentifier")
```

table 1 for all Marshallese and non-Marshallese

Baseline Characteristics	**Overall** N = 28,649[1]	**Marshallese** N = 873	**Non-Marshallese** N = 27776
Age			
Median (Min, Max)	41 (18, 102)	45 (18, 89)	41 (18, 102)
Sex			
Female	16,690 (58%)	577 (66%)	16,113 (58%)
Male	11,959 (42%)	296 (34%)	11,663 (42%)
Hypertension	6,906 (24%)	211 (24%)	6,695 (24%)
Pre-Diabetes or T2DM	4,506 (16%)	422 (48%)	4,084 (15%)
Hypertension and T2DM	2,806 (9.8%)	190 (22%)	2,616 (9.4%)
Clinic Location			
Maple	19,692 (69%)	432 (52%)	19,260 (69%)
Market	8,907 (31%)	392 (48%)	8,515 (31%)
Missing	50	49	1

Baseline Characteristics	Overall N = 28,649[1]	Marshallese N = 873	Non-Marshallese N = 27776
Housing Status			
Housed	20,628 (72%)	544 (62%)	20,084 (72%)
Doubling Up	5,585 (19%)	304 (35%)	5,281 (19%)
Permanent Supportive Housing	350 (1.2%)	7 (0.8%)	343 (1.2%)
Transitional	626 (2.2%)	4 (0.5%)	622 (2.2%)
Homeless Shelter	438 (1.5%)	5 (0.6%)	433 (1.6%)
Street	524 (1.8%)	2 (0.2%)	522 (1.9%)
Other	444 (1.5%)	6 (0.7%)	438 (1.6%)
Unknown	54 (0.2%)	1 (0.1%)	53 (0.2%)
Income Level			
Median (Min, Max)	114 (0, 25,568)	43 (0, 461)	116 (0, 25,568)
Missing	4,663	66	4,597
Risk Score			
Median (Min, Max)	1 (0, 22)	1 (0, 12)	1 (0, 22)
Missing	3,207	72	3,135

[1] n (%)