

Homework 9

Miles Tweed

2/4/2021

1. show the contingency table of the counts, using two categorical variables of your choice.

```
loans1 <-  
loans %>% select(purpose, grade) %>% filter(purpose != "", grade != "")  
  
table(loans1)
```

##	grade							
## purpose	A	B	C	D	E	F	G	
## car	6166	7385	6240	2891	987	275	69	
## credit_card	140508	176750	130553	48764	15775	3757	864	
## debt_consolidation	205916	361873	387312	200927	87262	26993	7594	
## educational	88	112	115	53	37	11	8	
## home_improvement	34528	44325	40963	19267	8120	2514	740	
## house	2579	3282	3541	2465	1370	597	302	
## major_purchase	11456	14638	13541	6875	2784	894	257	
## medical	4459	7306	8320	4777	1919	575	132	
## moving	1746	3402	4785	3276	1517	530	147	
## other	19571	34451	42119	26434	11556	3944	1365	
## renewable_energy	179	321	419	298	151	59	18	
## small_business	2963	4887	6671	5220	3009	1337	602	
## vacation	2419	4271	4978	2671	931	211	44	
## wedding	449	554	496	506	221	103	26	

2. Obtain the conditional proportions of one variable, given the categories of the other.

```
loans2 <-  
  loans1 %>% group_by(purpose) %>% count(grade) %>% mutate(Percent = 100*n/sum(n))  
  
loans2$purpose <-  
fct_recode(loans2$purpose, Car="car",  
  `Credit Card`="credit_card",  
  `Debt Consolidation`="debt_consolidation",  
  Educational="educational",  
  `Home Improvement`="home_improvement",  
  House="house",  
  `Major Purchases`="major_purchase",  
  Medical="medical",  
  Moving="moving",  
  Other="other",  
  `Renewable Energy`="renewable_energy",  
  `Small Business`="small_business",  
  Vacation="vacation",  
  Wedding="wedding")
```

```
head(loans2)
```

```
## # A tibble: 6 x 4
## # Groups:   purpose [1]
##   purpose grade      n Percent
##   <fct>   <chr> <int>   <dbl>
## 1 Car      A      6166   25.7
## 2 Car      B      7385   30.8
## 3 Car      C      6240   26.0
## 4 Car      D      2891   12.0
## 5 Car      E       987    4.11
## 6 Car      F       275    1.15
```

3. Construct an appropriate graph to compare the conditional proportions.

```
loans2 %>% ggplot(aes(y = Percent, x = grade, fill = grade)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(vars(purpose)) +
  scale_fill_brewer(palette = "Set3", direction = -1) +
  guides(fill = 'none') +
  theme_minimal() +
  labs(title = "Distribution of Loan Grades by Loan Purpose") + xlab("Grade")
```

