

Project 1

Sophia Telles & Samantha Driscoll

October 20, 2017

```
load(file="providerspokane.rda")
```

```
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(tidyr)
library(RColorBrewer)
```

Spokane Providers

Intro

Medical providers in the Spokane area can be analyzed as a function of Gender of the provider, type of service provided, and place of service. These factors, when analyzed as functions of variables such as Medicare coverage, tell us a lot about what services in the medical field are used the most frequently, cost the most money, and have the most Medicare coverage.

Question 1

Are there differences in the number of services, distinct beneficiary per day services, average medicare allowed, charged, and paid amount, and medicare standardized amount differ as a function of Gender, the Provider Type, and Place of Service?

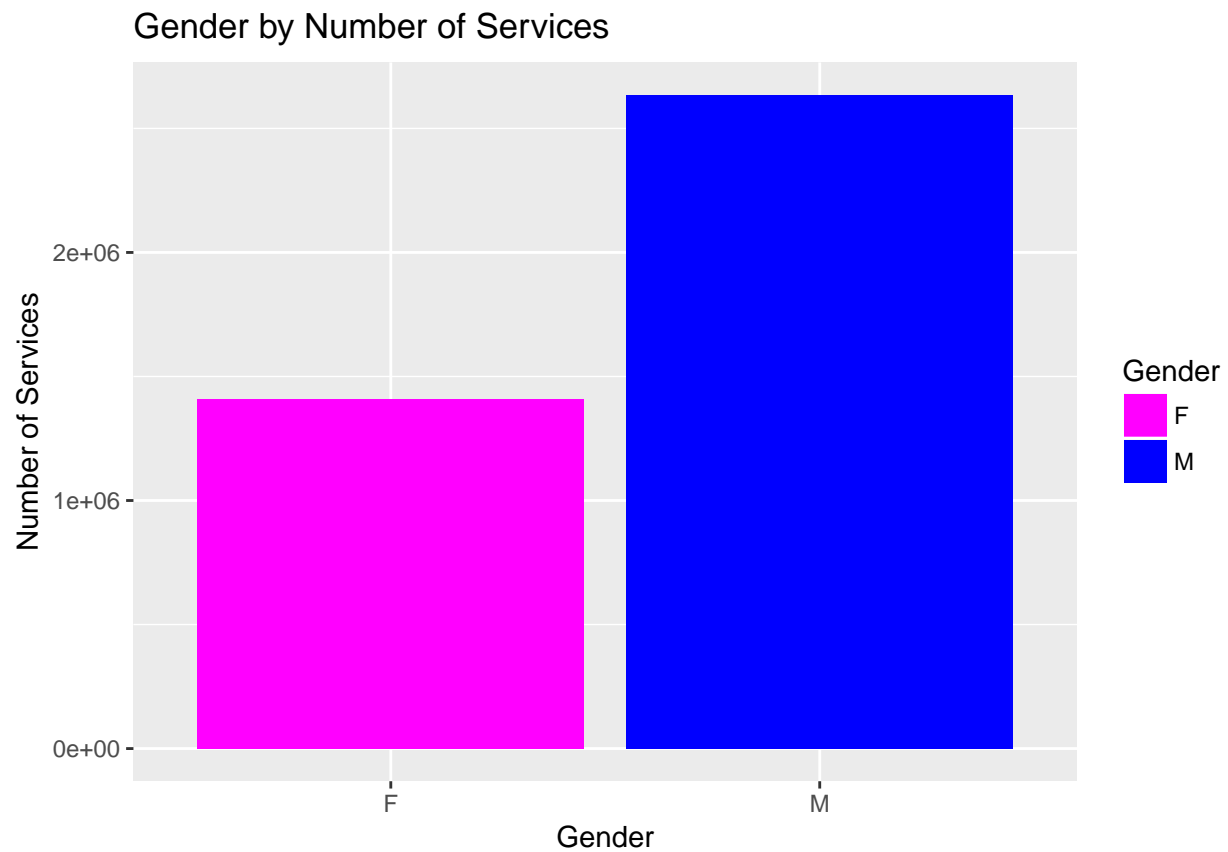
Gender

Number of Services and Gender

```
ServicesG=providerspokane%>%group_by(Gender=Gender.of.the.Provider)%>%summarize(NumberofServices =sum(N
ServicesG[ServicesG==""] <-NA
ServicesG = ServicesG[complete.cases(ServicesG),]
```

We have some service providers that did not specify Gender, so we discovered a way to eliminate this variable of the data.

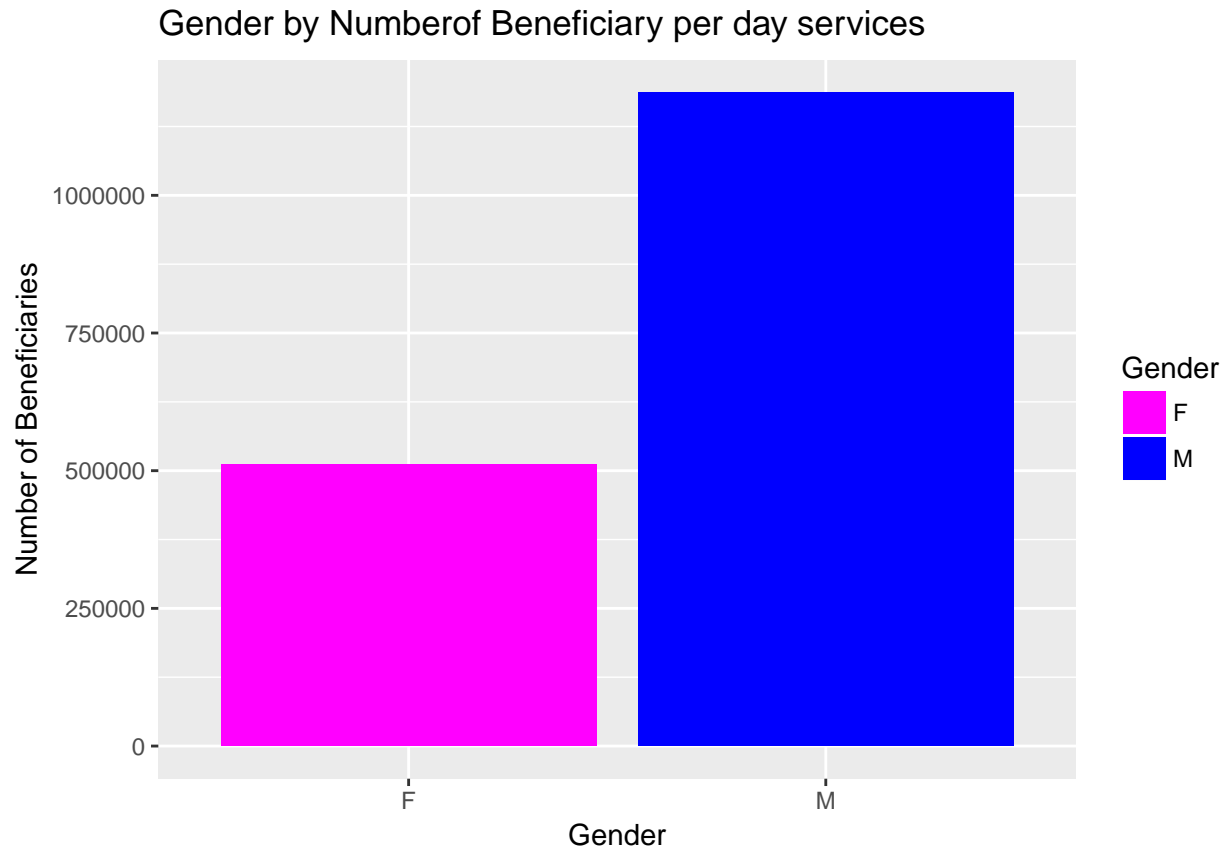
```
ggplot(ServicesG,aes(Gender,NumberofServices,fill=Gender))+geom_bar(stat="identity")+scale_fill_manual(
```



Number of Beneficiary per day Services and Gender

```
BeneficiaryG=providerspokane%>%group_by(Gender = Gender.of.the.Provider)%>%summarize(NumberofBeneficiary=
NumberofBeneficiary)
BeneficiaryG[BeneficiaryG==""] <-NA
BeneficiaryG = BeneficiaryG[complete.cases(BeneficiaryG),]
```

```
ggplot(BeneficiaryG,aes(Gender,NumberofBeneficiary,fill=Gender))+geom_bar(stat="identity")+scale_fill_m
```

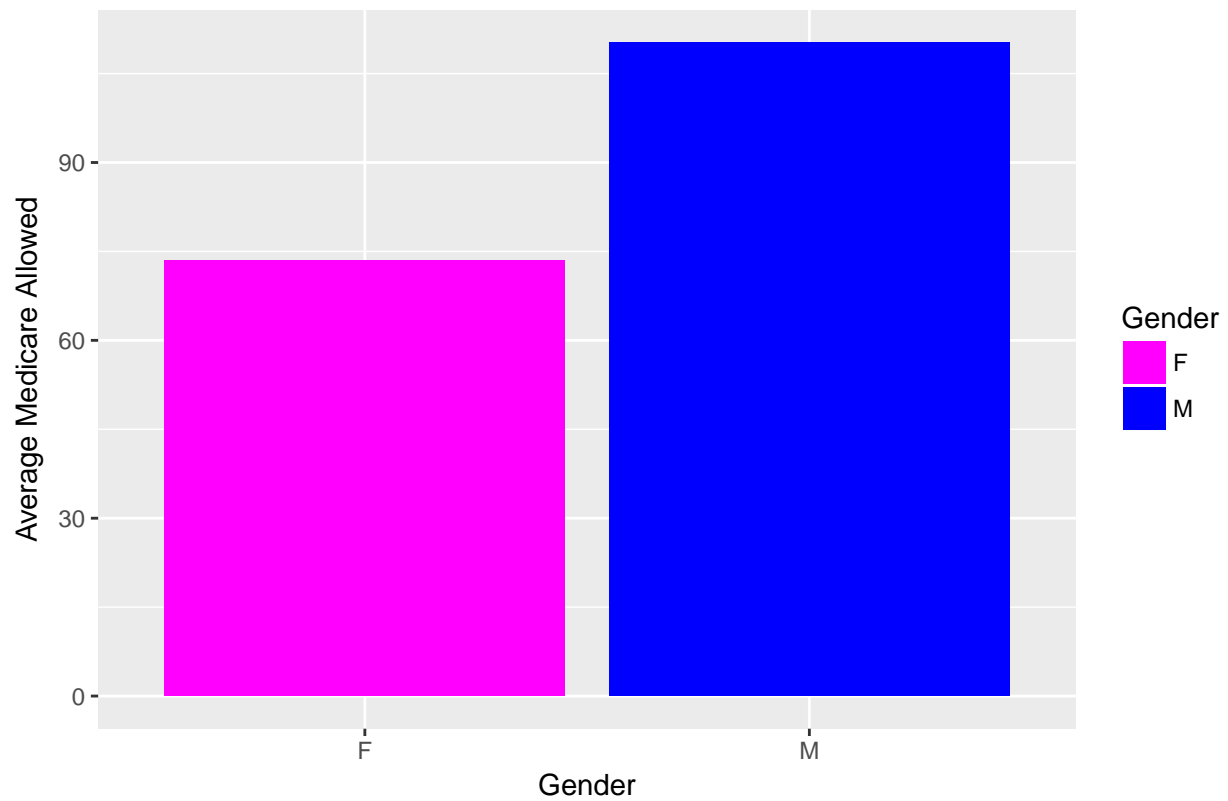


Average Medicare Allowed and Gender

```
AllowedG=providerspokane%>%group_by(Gender=Gender.of.the.Provider)%>%summarize(AllowedAmount=mean(Aver
AllowedG[AllowedG==""] <-NA
AllowedG = AllowedG[complete.cases(AllowedG),]
```

```
ggplot(AllowedG,aes(Gender,AllowedAmount,fill=Gender))+geom_bar(stat="identity")+scale_fill_manual(val
```

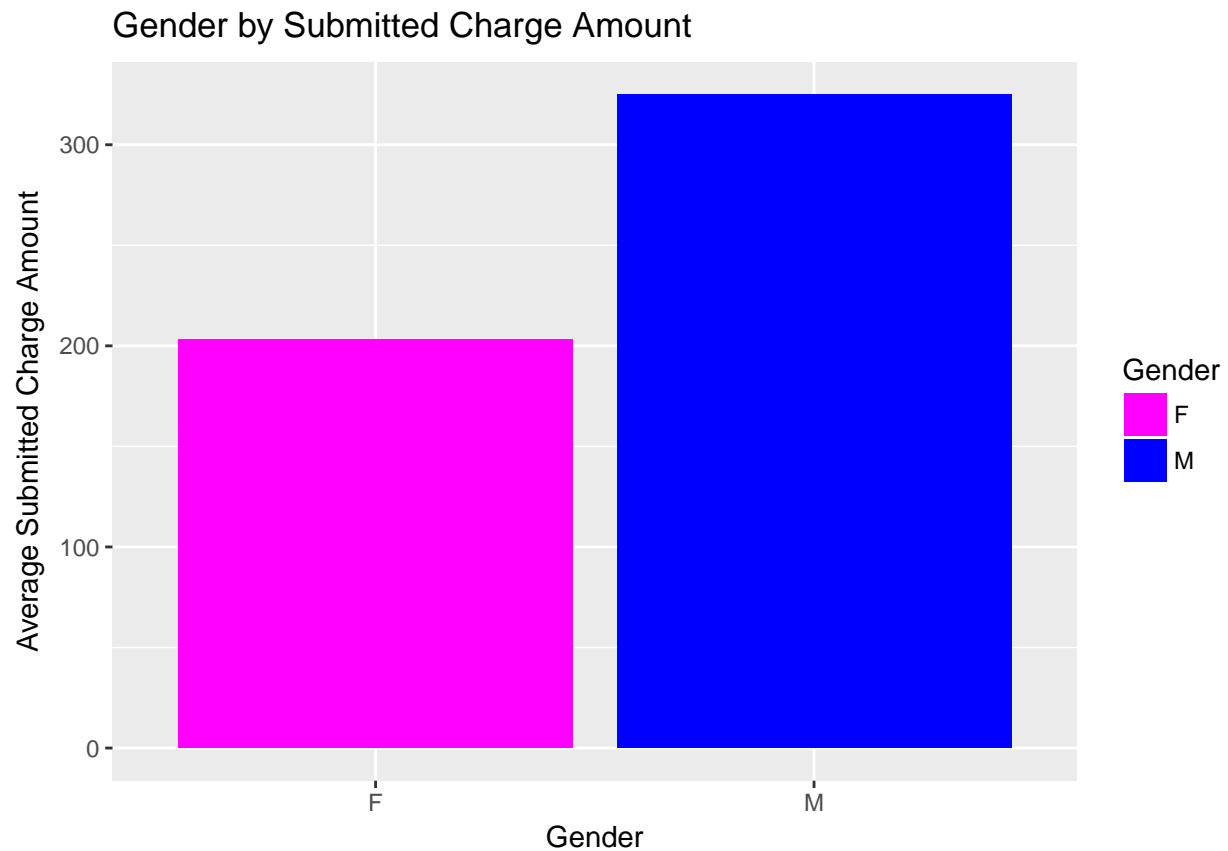
Gender by Average Medicare Allowed



Average Submitted Charge Amount and Gender

```
ChargedG=providerspokane%>%group_by(Gender=Gender.of.the.Provider)%>%summarize(ChargedAmmount=mean(AverageSubmittedChargeAmount))
ChargedG[ChargedG==""] <-NA
ChargedG = ChargedG[complete.cases(ChargedG),]

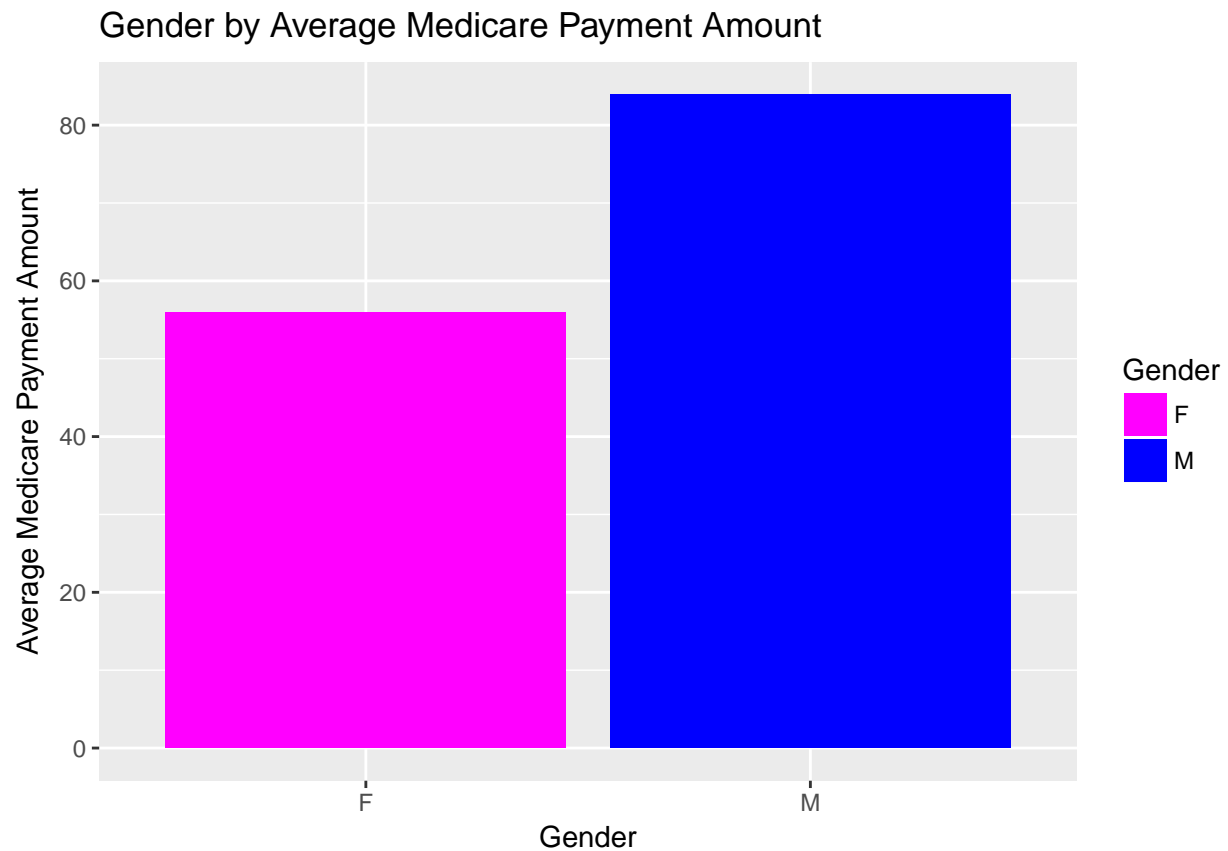
ggplot(ChargedG,aes(Gender,ChargedAmmount,fill=Gender))+geom_bar(stat="identity")+scale_fill_manual(values=c("F"="#FF00FF","M"="#0000FF"))
```



Average Medicare Payment Amount and Gender

```
PaymentG=providerspokane%>%group_by(Gender=Gender.of.the.Provider)%>%summarize(PaymentAmount=mean(AverageMedicarePaymentAmount))
PaymentG[PaymentG==""] <-NA
PaymentG = PaymentG[complete.cases(PaymentG),]

ggplot(PaymentG,aes(Gender,PaymentAmount,fill=Gender))+geom_bar(stat="identity")+scale_fill_manual(values=c("F"="#FF00FF","M"="#0000FF"))
```

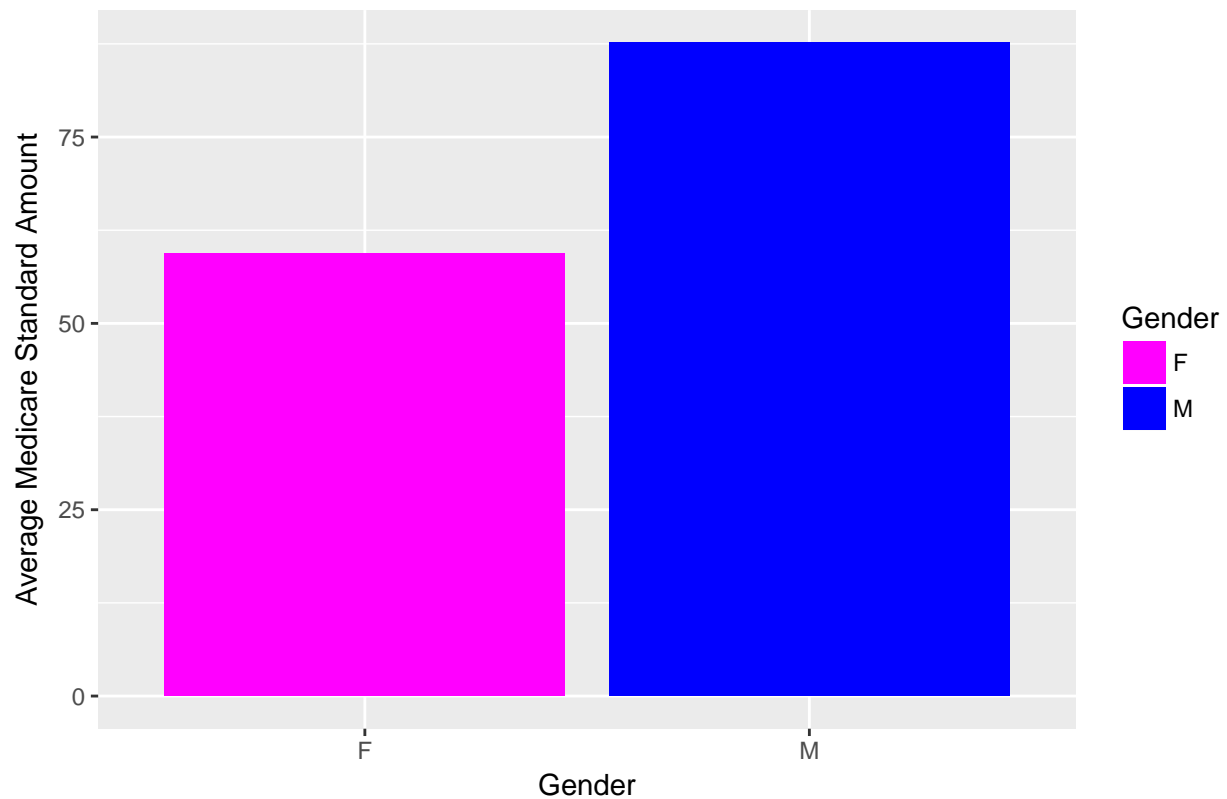


Average Medicare Standardized Amount and Gender

```
StandardG=providerspokane%>%group_by(Gender=Gender.of.the.Provider)%>%summarize(StandardAmmount=mean(AverageMedicarePaymentAmount))
StandardG[StandardG==""] <-NA
StandardG = StandardG[complete.cases(StandardG),]

ggplot(StandardG,aes(Gender,StandardAmmount,fill=Gender))+geom_bar(stat="identity")+scale_fill_manual(values=c("F"="#FF00FF","M"="#0000FF"))
```

Gender by Average Medicare Standardized Amount



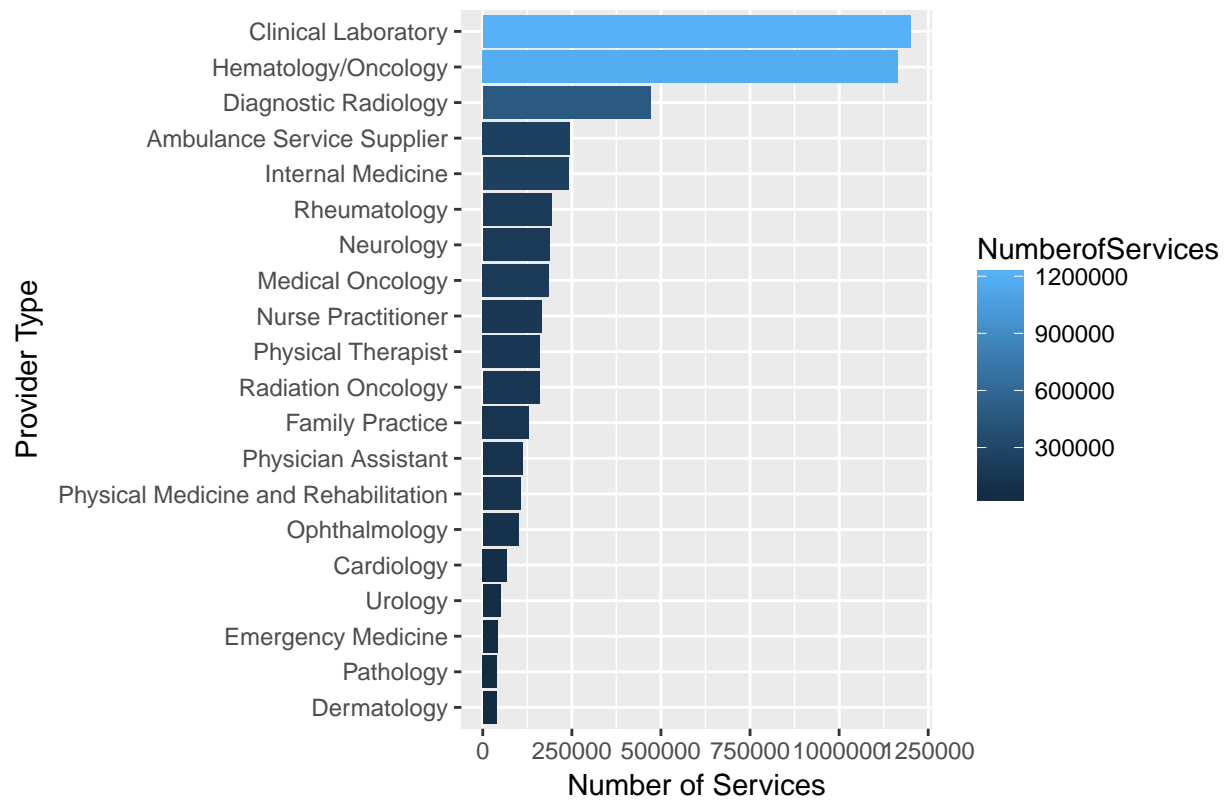
Provider Type

Number of Services and Provider Type

```
ServicesP=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(NumberofServices =sum(Number.of
ServicesP <- ServicesP[order(-ServicesP$NumberofServices),]
ServicesP<-ServicesP[1:20,]
ServicesPs=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(NumberofServices =sum(Number.of
ServicesPs <- ServicesPs[order(-ServicesPs$NumberofServices),]
ServicesPs <-ServicesPs[21:45,]
ServicesPt=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(NumberofServices =sum(Number.of
ServicesPt <- ServicesPt[order(-ServicesPt$NumberofServices),]
ServicesPt <-ServicesPt[46:70,]
```

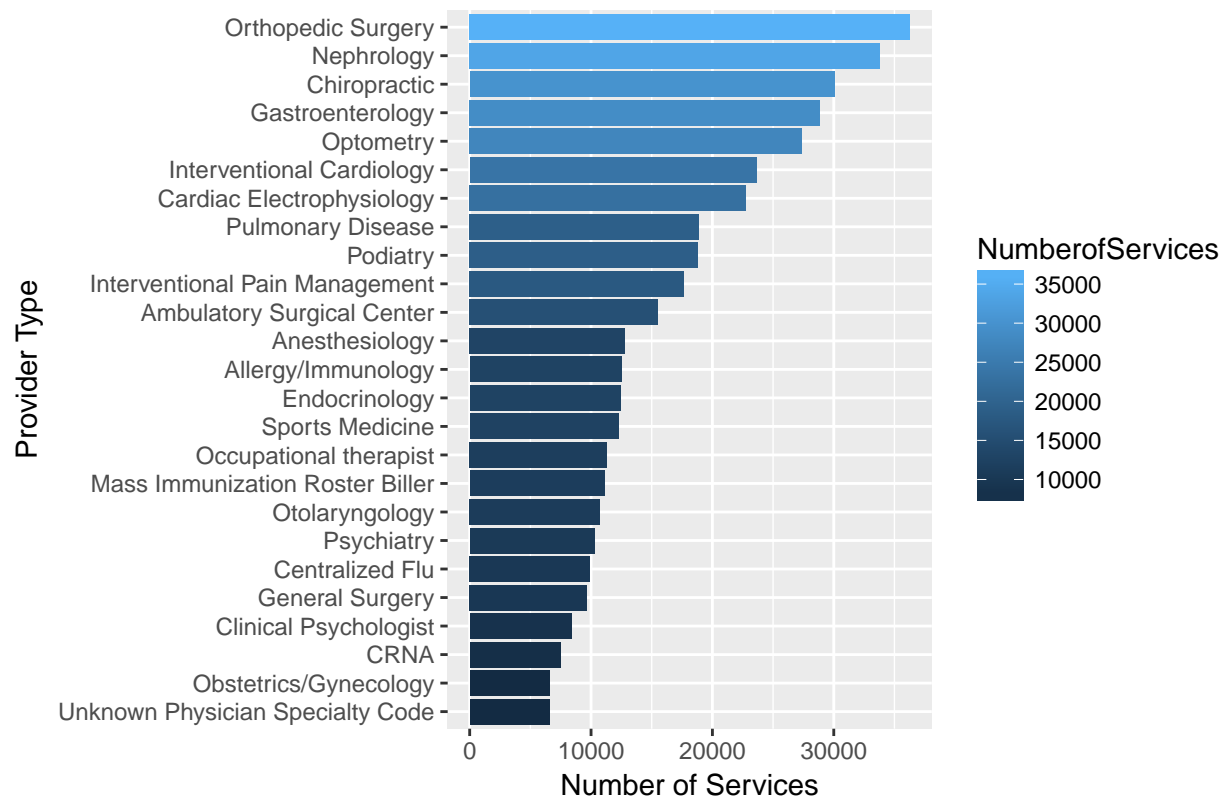
```
ggplot(ServicesP,aes(reorder(Provider,NumberofServices),NumberofServices,fill=NumberofServices))+geom_bar()
```

Provider Type 1–20 categorized by Number of Serv



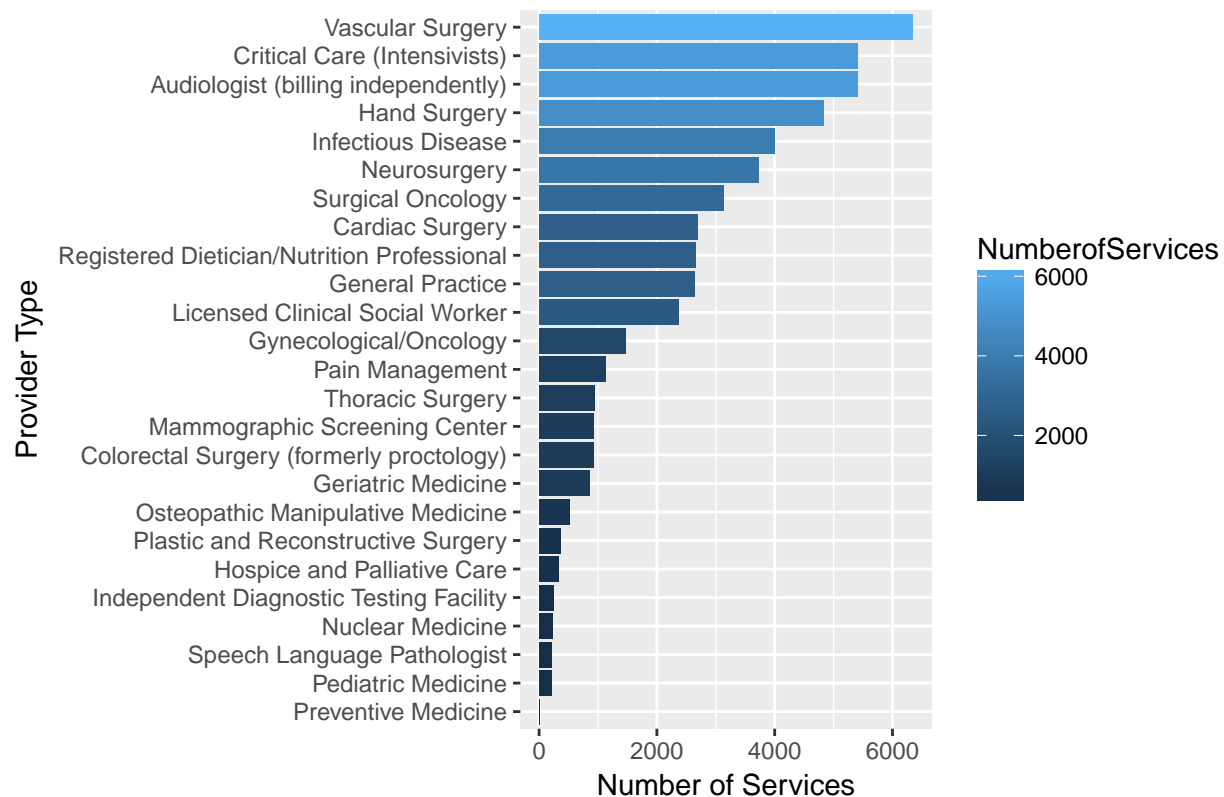
```
ggplot(ServicesPs, aes(reorder(Provider, NumberofServices), NumberofServices, fill=NumberofServices)) + geom_bar()
```


Provider Type 21–46 categorized by Number of Serv



```
ggplot(ServicesPt, aes(reorder(Provider, NumberofServices), NumberofServices, fill=NumberofServices)) + geom_bar()
```

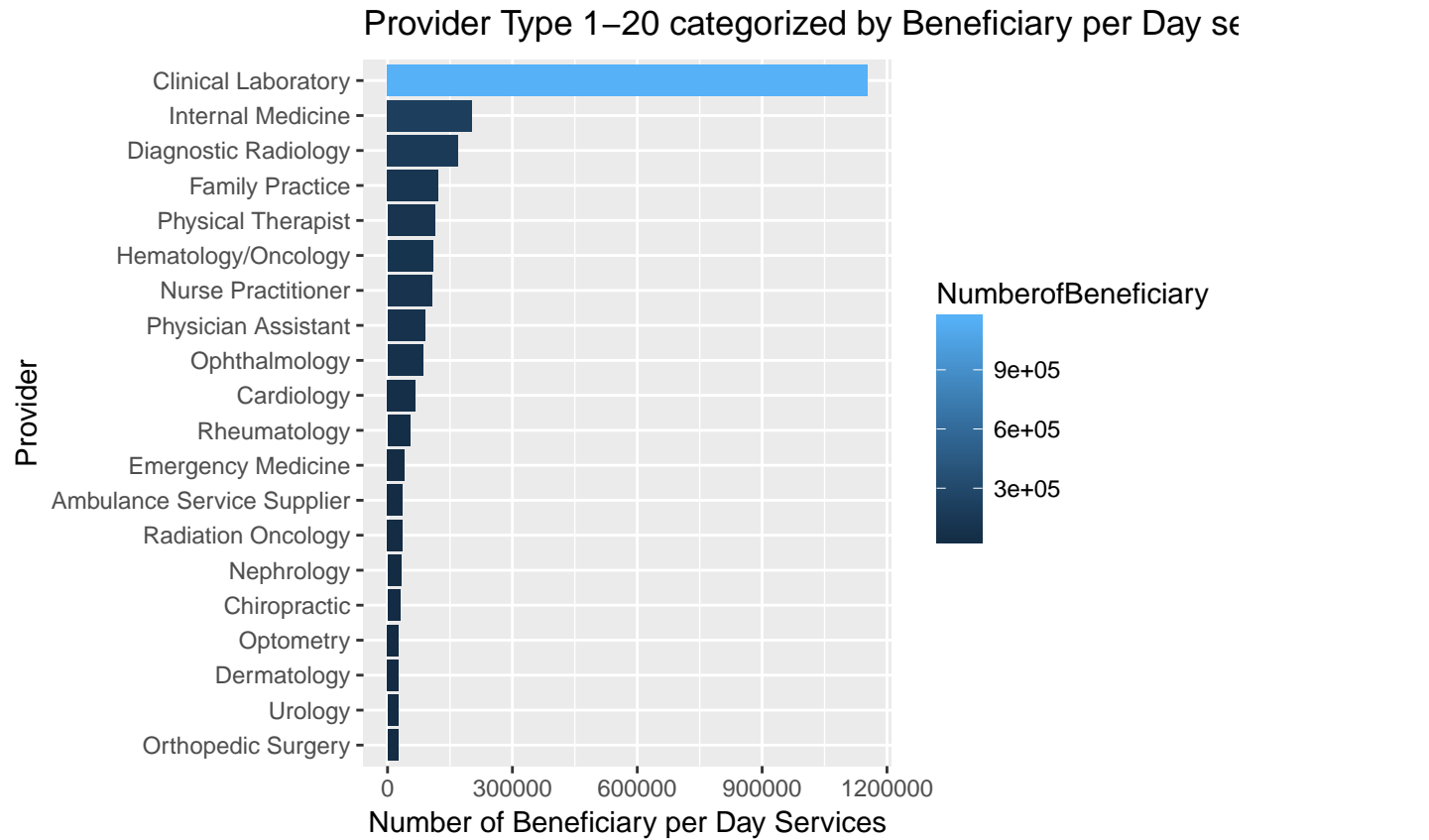
Provider Type 46–70 categorized by Number of



Number of Beneficiary per day Services and Provider Type

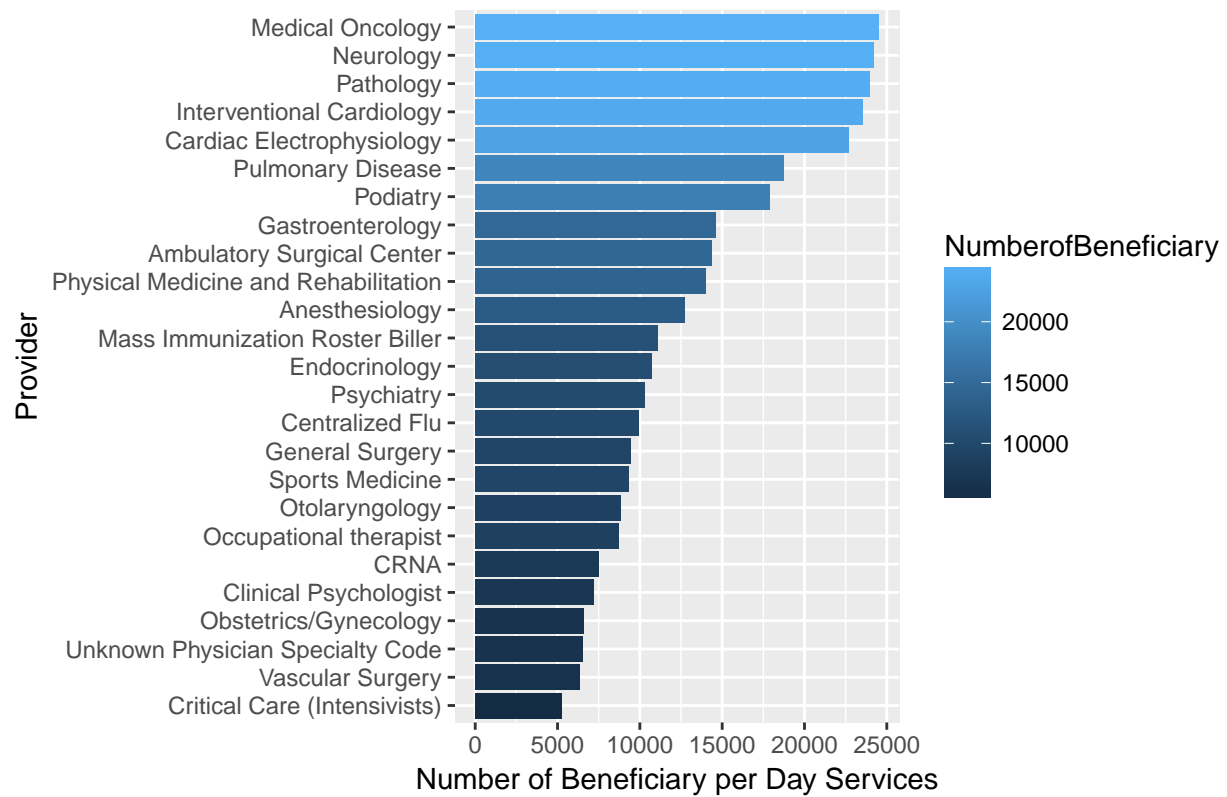
```
BeneficiaryP=providerspokane%>%group_by(Provider =Provider.Type)%>%summarize(NumberofBeneficiary=sum(Nu
BeneficiaryP <- BeneficiaryP[order(-BeneficiaryP$NumberofBeneficiary),]
BeneficiaryP <- BeneficiaryP[1:20,]
BeneficiaryPs=providerspokane%>%group_by(Provider =Provider.Type)%>%summarize(NumberofBeneficiary=sum(Nu
BeneficiaryPs <- BeneficiaryPs[order(-BeneficiaryPs$NumberofBeneficiary),]
BeneficiaryPs <- BeneficiaryPs[21:45,]
BeneficiaryPt=providerspokane%>%group_by(Provider =Provider.Type)%>%summarize(NumberofBeneficiary=sum(Nu
BeneficiaryPt <- BeneficiaryPt[order(-BeneficiaryPt$NumberofBeneficiary),]
BeneficiaryPt <- BeneficiaryPt[46:70,]

ggplot(BeneficiaryP,aes(reorder(Provider,NumberofBeneficiary),NumberofBeneficiary, fill=NumberofBenefic.
```



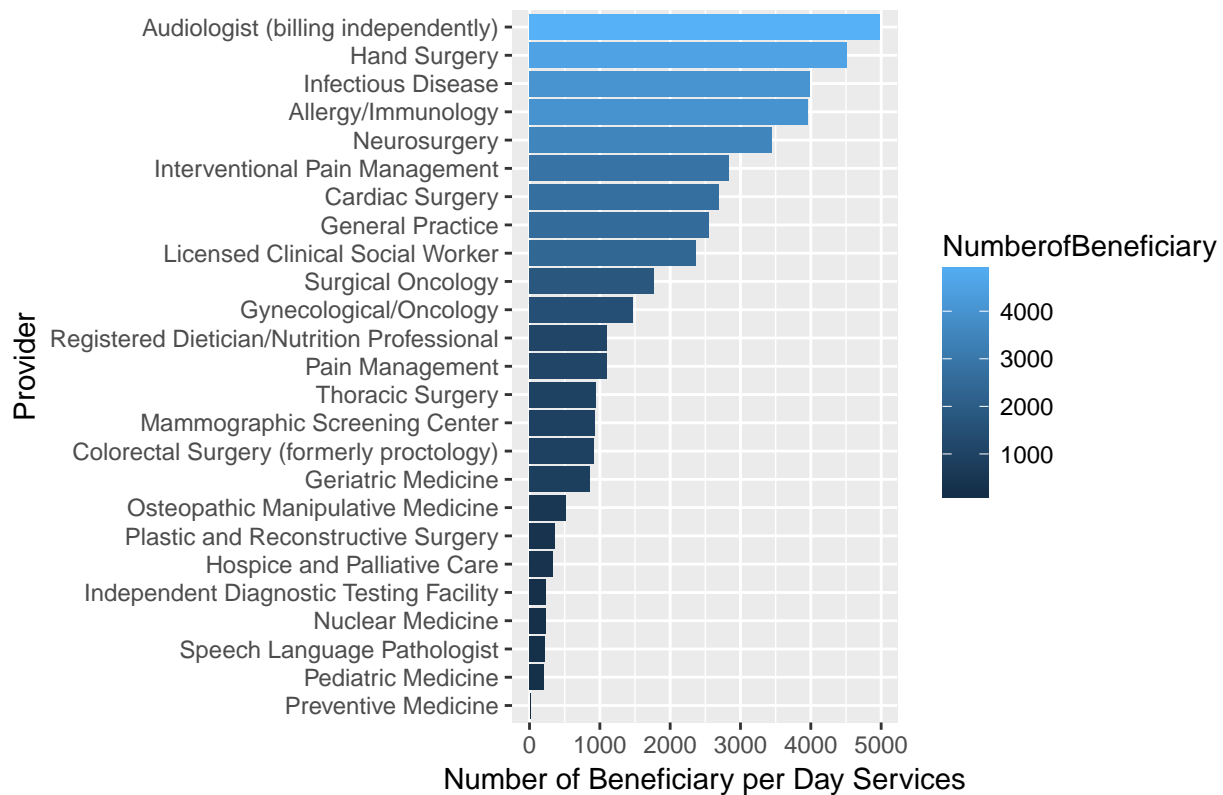
```
ggplot(BeneficiaryPs,aes(reorder(Provider,NumberofBeneficiary),NumberofBeneficiary, fill=NumberofBenefi
```

Provider Type 21–45 categorized by Beneficiary per



```
ggplot(BeneficiaryPt, aes(reorder(Provider, NumberofBeneficiary), NumberofBeneficiary, fill=NumberofBenefi
```

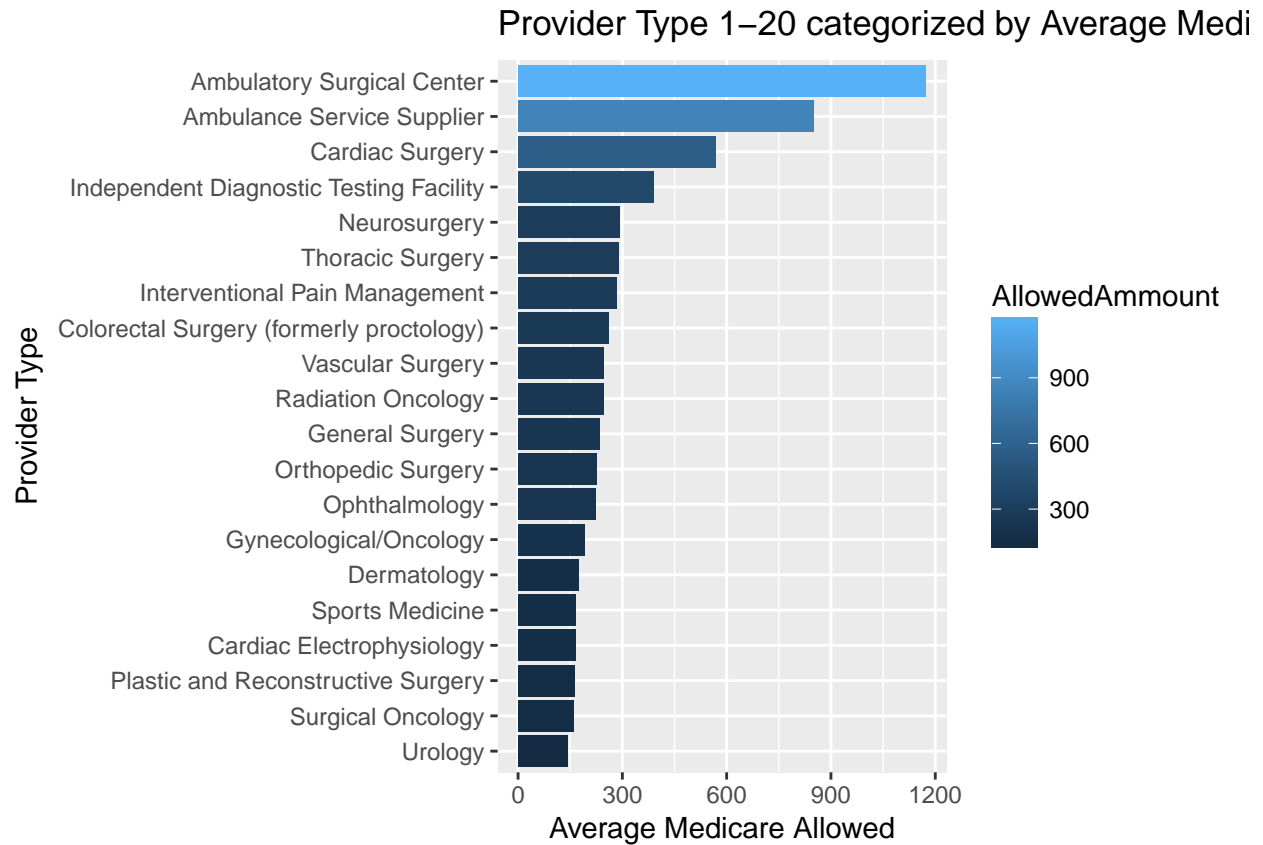
Provider Type 46–70 categorized by Beneficiary



Average Medicare Allowed and Provider Type

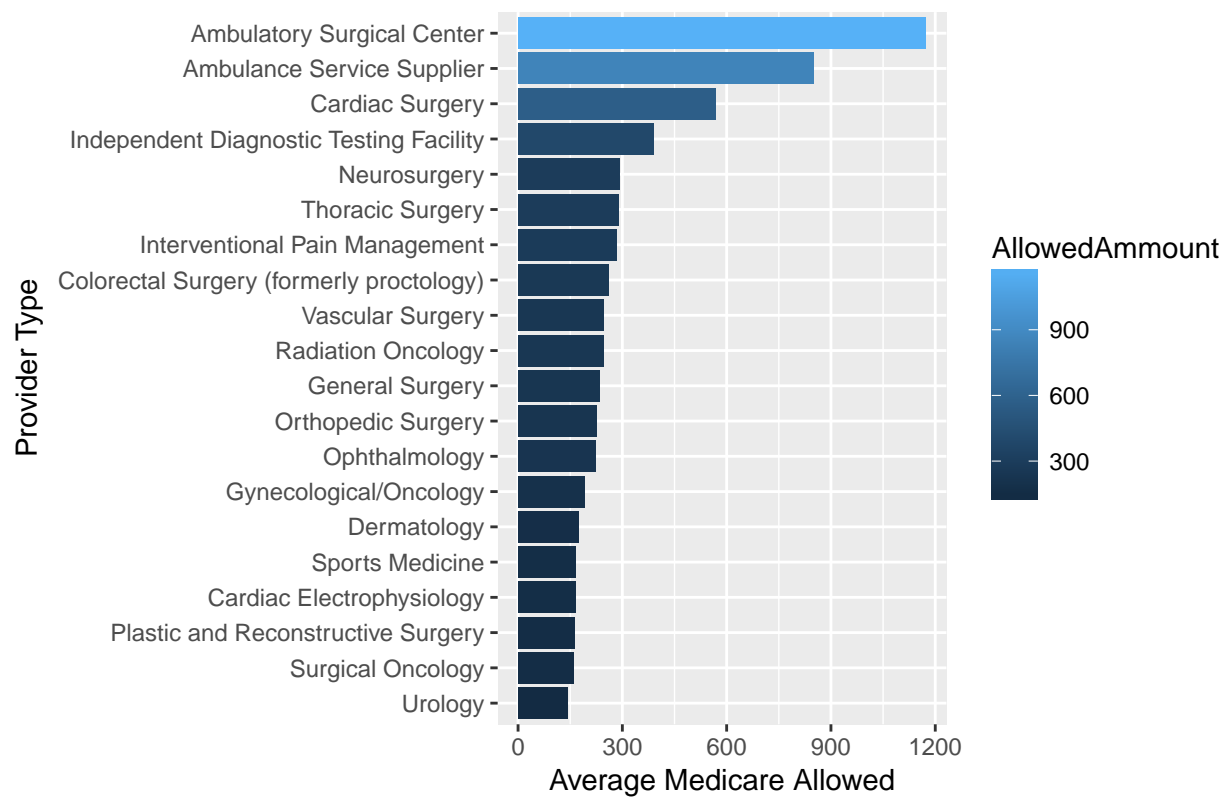
```
AllowedP=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(AllowedAmount=mean(Average.Medicare.Allowable.Amount))
AllowedP<- AllowedP[order(-AllowedP$AllowedAmount),]
AllowedP<- AllowedP[1:20,]
AllowedPs=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(AllowedAmount=mean(Average.Medicare.Allowable.Amount))
AllowedPs<- AllowedPs[order(-AllowedPs$AllowedAmount),]
AllowedPs<- AllowedPs[1:20,]
AllowedPt=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(AllowedAmount=mean(Average.Medicare.Allowable.Amount))
AllowedPt<- AllowedPt[order(-AllowedPt$AllowedAmount),]
AllowedPt<- AllowedPt[21:45,]

ggplot(AllowedP,aes(reorder(Provider,AllowedAmount),AllowedAmount,fill=AllowedAmount))+geom_bar(stat="sum")
```

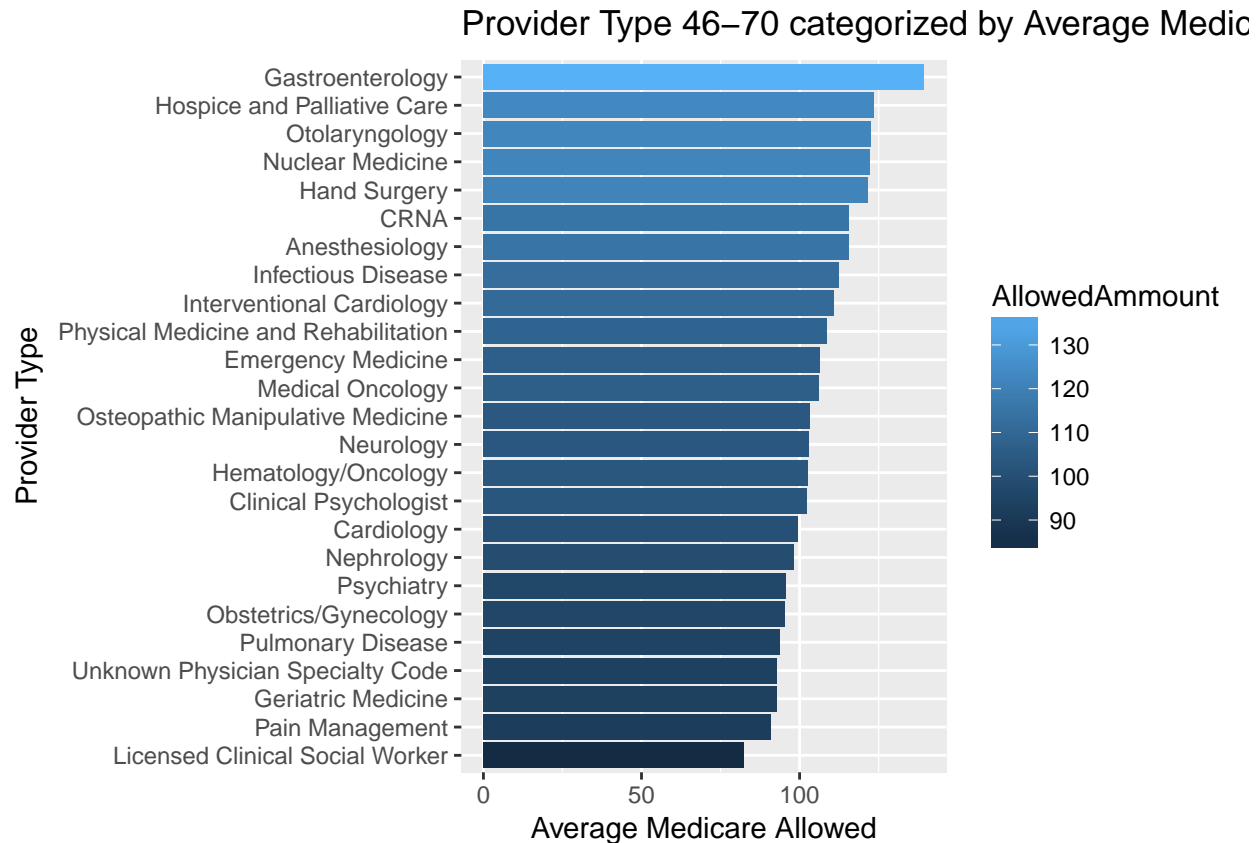


```
ggplot(AllowedPs, aes(reorder(Provider, AllowedAmmount), AllowedAmmount, fill=AllowedAmmount)) + geom_bar(Sta
```

Provider Type 21–46 categorized by Average Medicare Allowed



```
ggplot(AllowedPt, aes(reorder(Provider, AllowedAmount), AllowedAmount, fill=AllowedAmount)) + geom_bar(stat="identity")
```

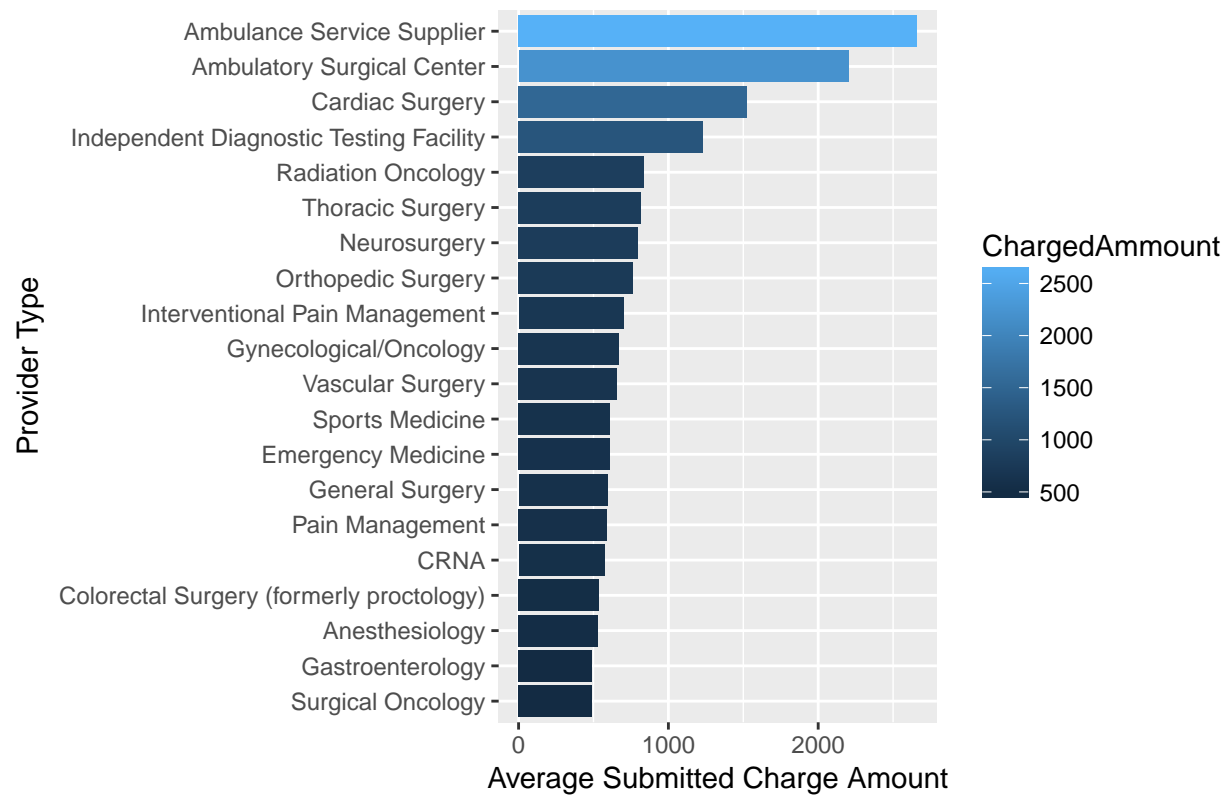


Average Submitted Charge Amount and Provider Type

```
ChargedP=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(ChargedAmount=mean(Average.Submitted.Charge.Amount))
ChargedP<- ChargedP[order(-ChargedP$ChargedAmount),]
ChargedP<- ChargedP[1:20,]
ChargedPs=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(ChargedAmount=mean(Average.Submitted.Charge.Amount))
ChargedPs<- ChargedPs[order(-ChargedPs$ChargedAmount),]
ChargedPs<- ChargedPs[21:45,]
ChargedPt=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(ChargedAmount=mean(Average.Submitted.Charge.Amount))
ChargedPt<- ChargedPt[order(-ChargedPt$ChargedAmount),]
ChargedPt<- ChargedPt[46:70,]

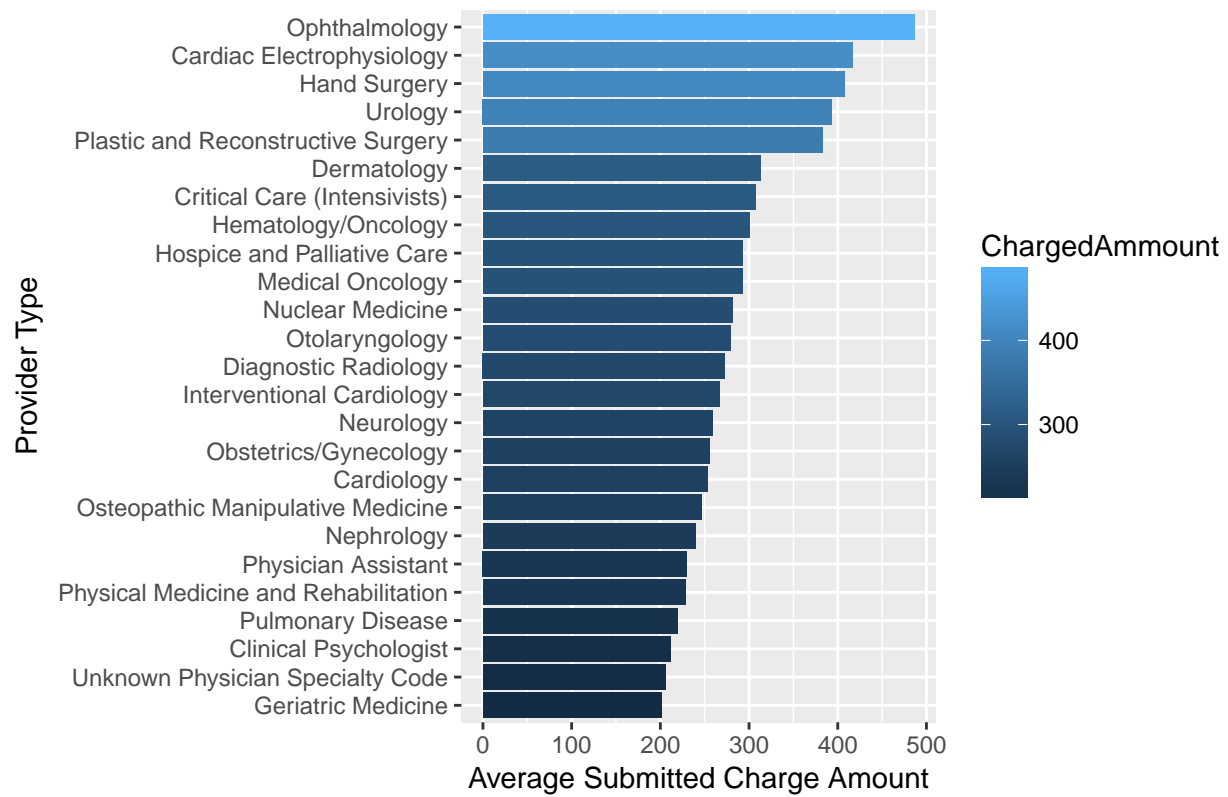
ggplot(ChargedP,aes(reorder(Provider,ChargedAmount),ChargedAmount,fill=ChargedAmount))+geom_bar(stat="summary",position="dodge")
```


Provider Type 1–20 categorized by Average Subr



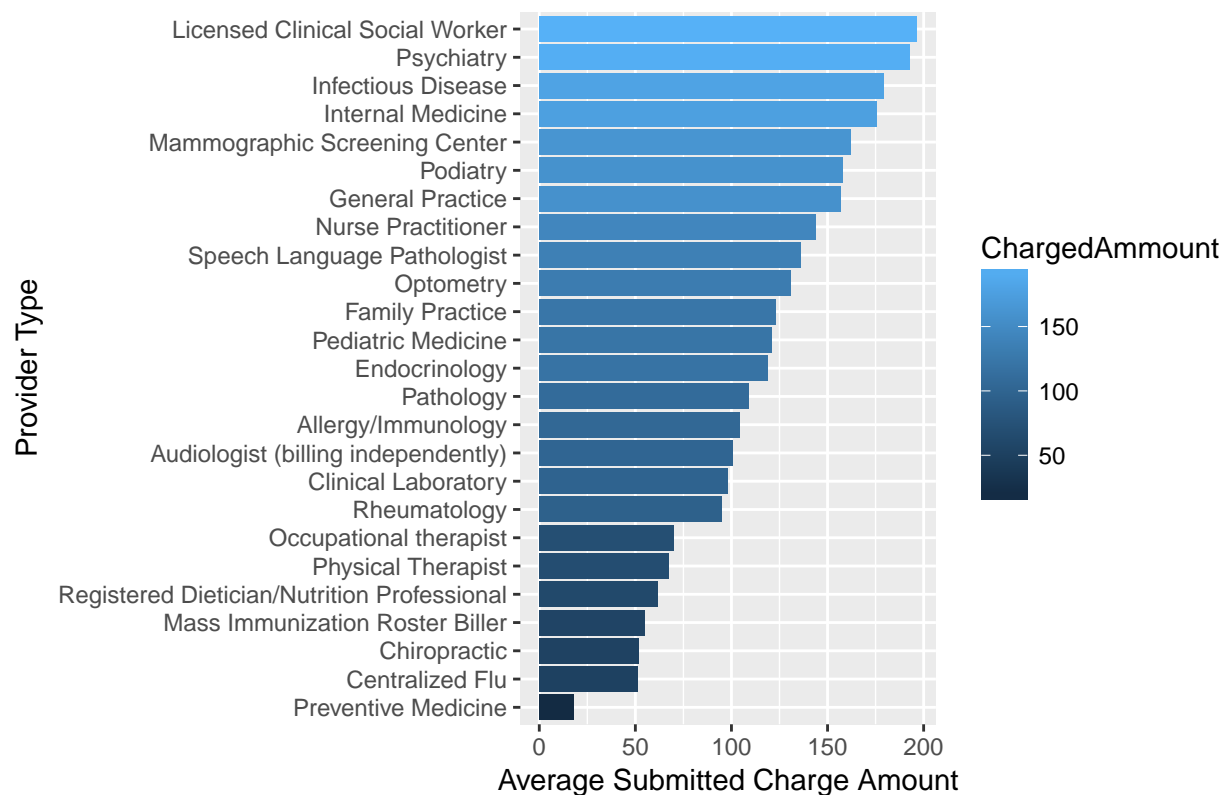
```
ggplot(ChargedPs, aes(reorder(Provider, ChargedAmmount), ChargedAmmount, fill=ChargedAmmount)) + geom_bar(Sta
```

Provider Type 21–45 categorized by Average Subm



```
ggplot(ChargedPt, aes(reorder(Provider, ChargedAmmount), ChargedAmmount, fill=ChargedAmmount)) + geom_bar(Sta
```

Provider Type 46–70 categorized by Average Submitted Charge Amount



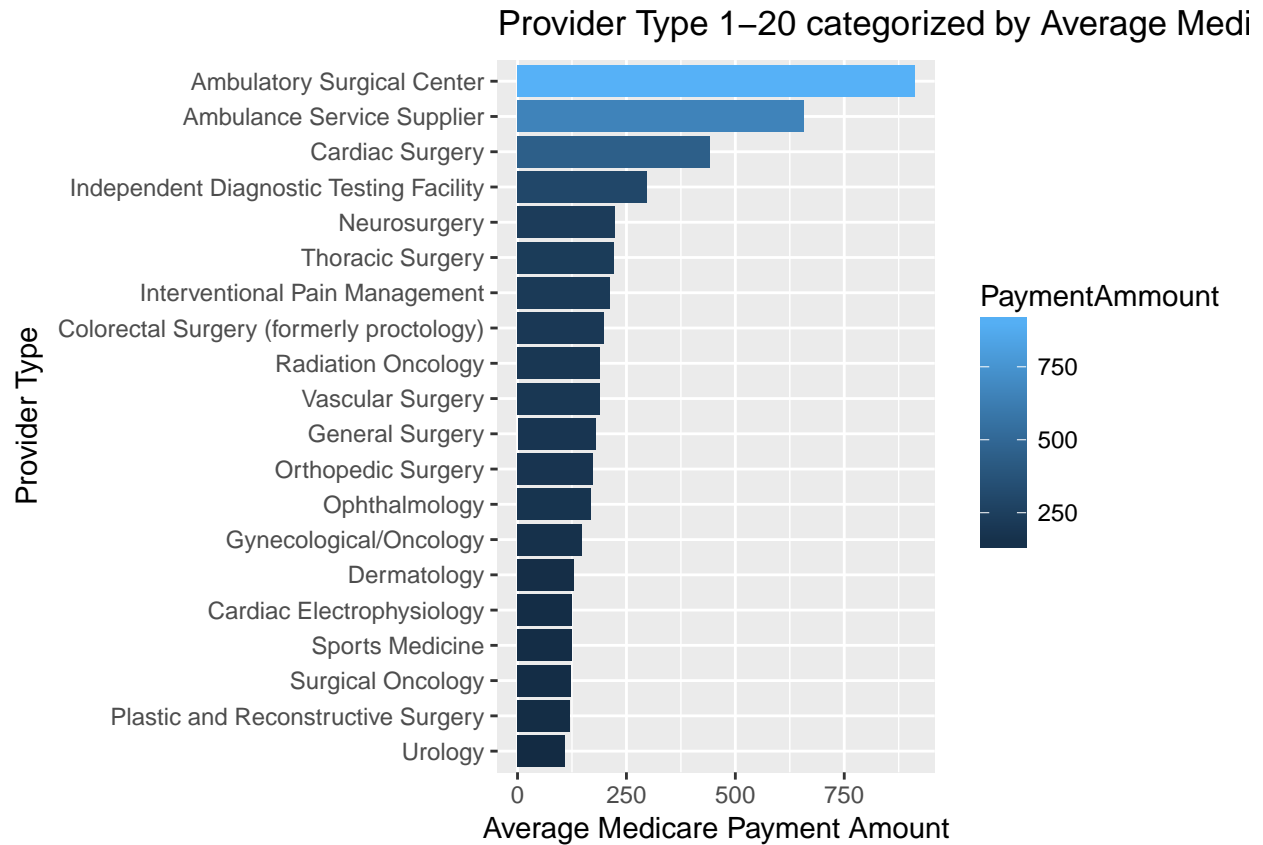
Average Medicare Payment Amount and Provider Type

```

PaymentP=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(PaymentAmount=mean(Average.Medicare.Payment))
PaymentP<- PaymentP[order(-PaymentP$PaymentAmount),]
PaymentP<- PaymentP[1:20,]
PaymentPs=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(PaymentAmount=mean(Average.Medicare.Payment))
PaymentPs<- PaymentPs[order(-PaymentPs$PaymentAmount),]
PaymentPs<- PaymentPs[21:45,]
PaymentPt=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(PaymentAmount=mean(Average.Medicare.Payment))
PaymentPt<- PaymentPt[order(-PaymentPt$PaymentAmount),]
PaymentPt<- PaymentPt[46:70,]

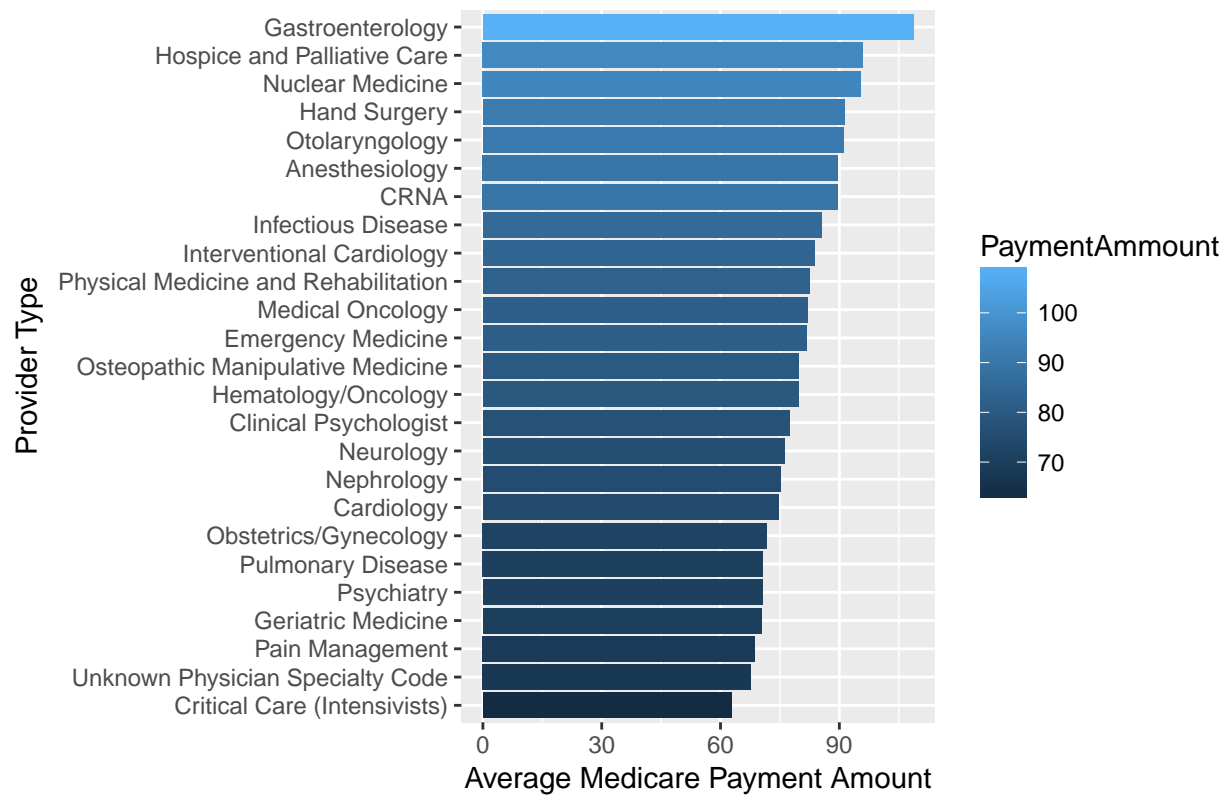
ggplot(PaymentP,aes(reorder(Provider,PaymentAmount),PaymentAmount,fill=PaymentAmount))+geom_bar(stat="summary")

```



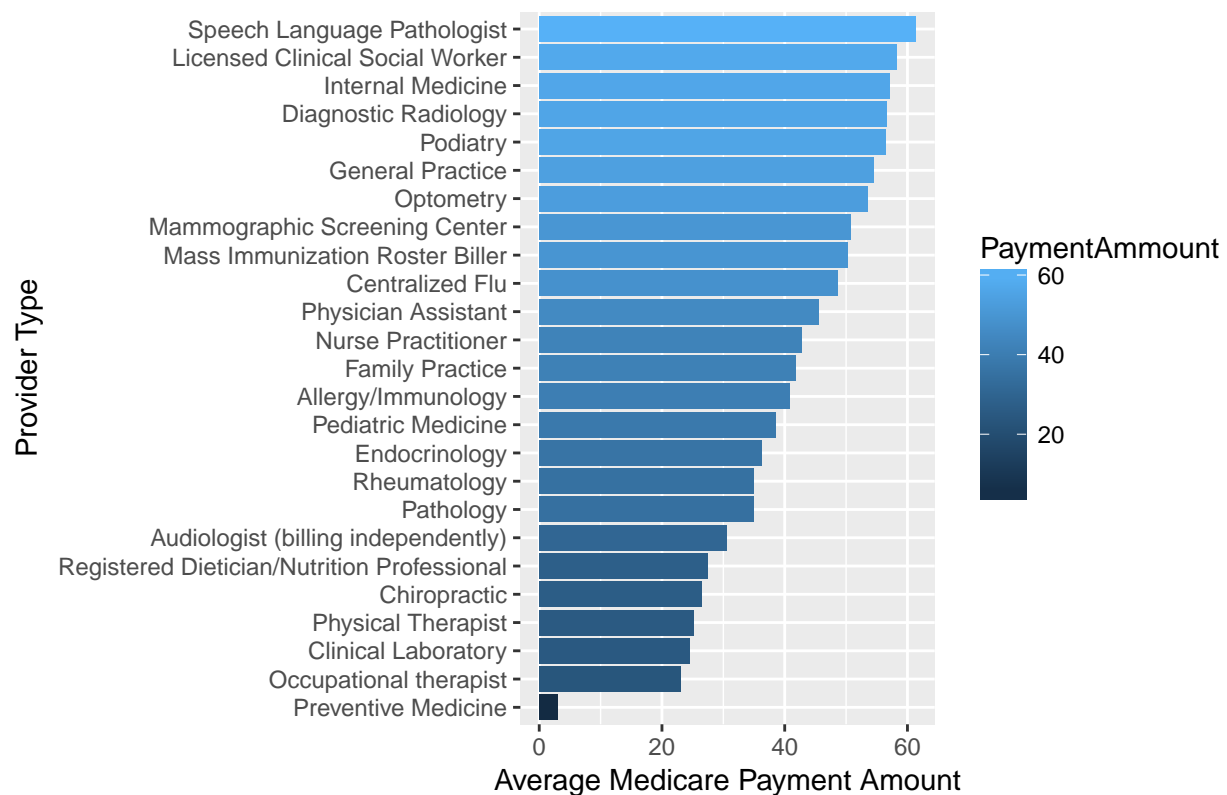
```
ggplot(PaymentPs, aes(reorder(Provider, PaymentAmmount), PaymentAmmount, fill=PaymentAmmount)) + geom_bar(
```

Provider Type 21–45 categorized by Average Medicare



```
ggplot(PaymentPt, aes(reorder(Provider, PaymentAmount), PaymentAmount, fill=PaymentAmount)) + geom_bar(stat="identity")
```

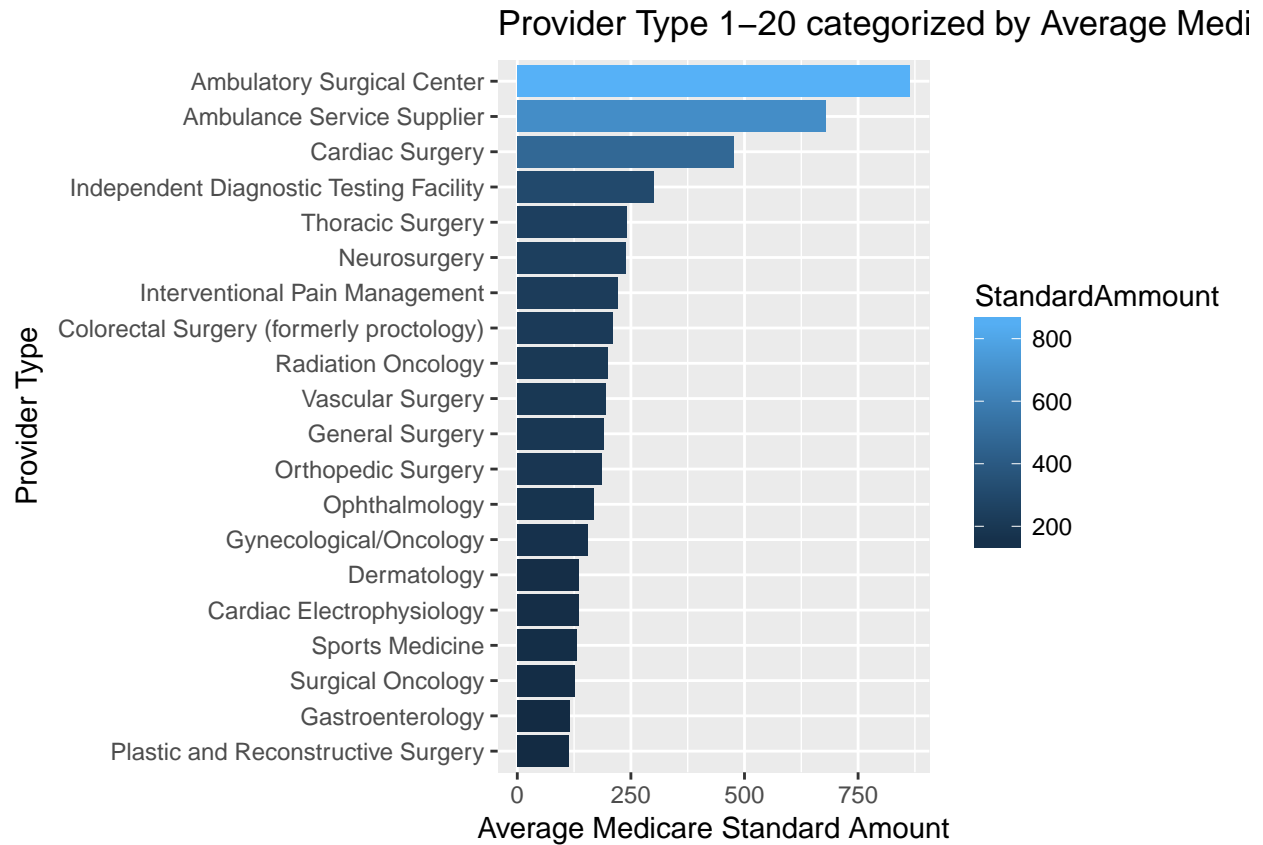
Provider Type 46–70 categorized by Average M



Average Medicare Standardized Amount and Provider Type

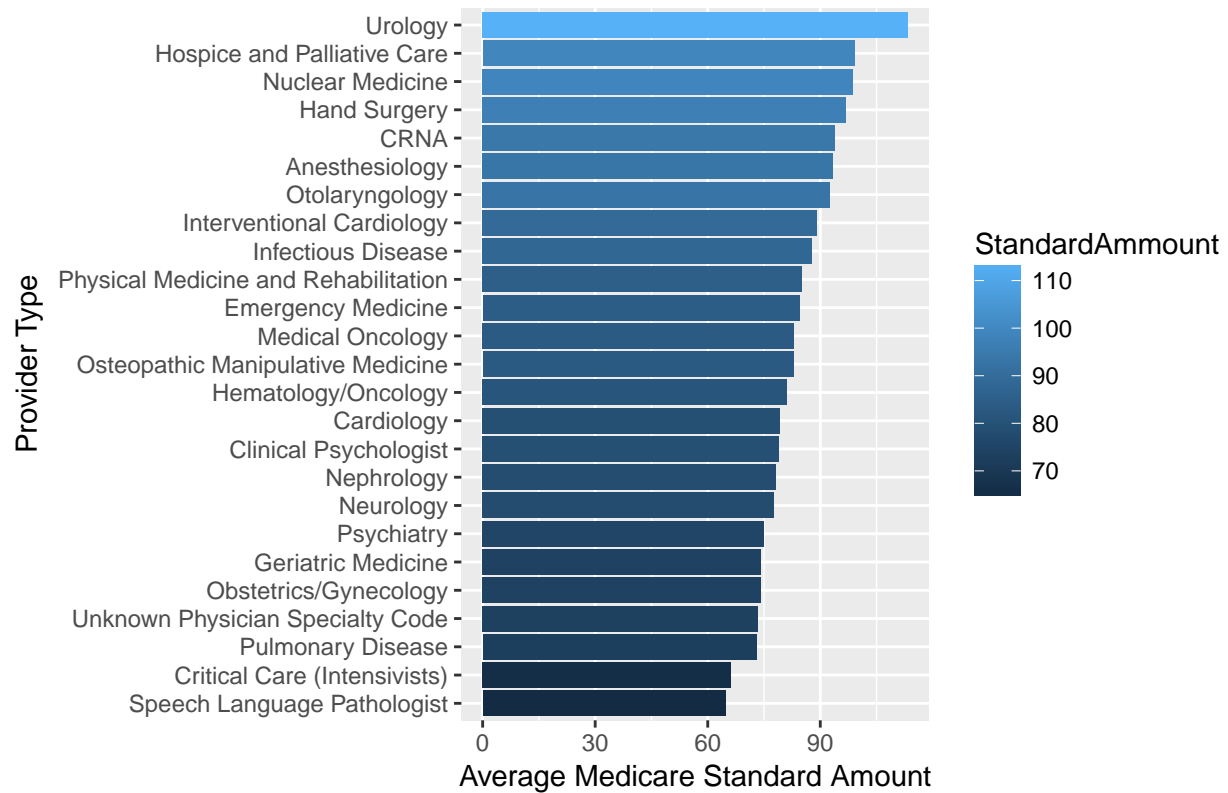
```
StandardP=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(StandardAmount=mean(Average.Medicare.Payment.Amount))
StandardP <- StandardP[order(-StandardP$StandardAmount),]
StandardP <- StandardP[1:20,]
StandardPs=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(StandardAmount=mean(Average.Medicare.Payment.Amount))
StandardPs <- StandardPs[order(-StandardPs$StandardAmount),]
StandardPs <- StandardPs[21:45,]
StandardPt=providerspokane%>%group_by(Provider=Provider.Type)%>%summarize(StandardAmount=mean(Average.Medicare.Payment.Amount))
StandardPt <- StandardPt[order(-StandardPt$StandardAmount),]
StandardPt <- StandardPt[46:70,]

ggplot(StandardP,aes(reorder(Provider,StandardAmount),StandardAmount,fill=StandardAmount))+geom_bar()
```

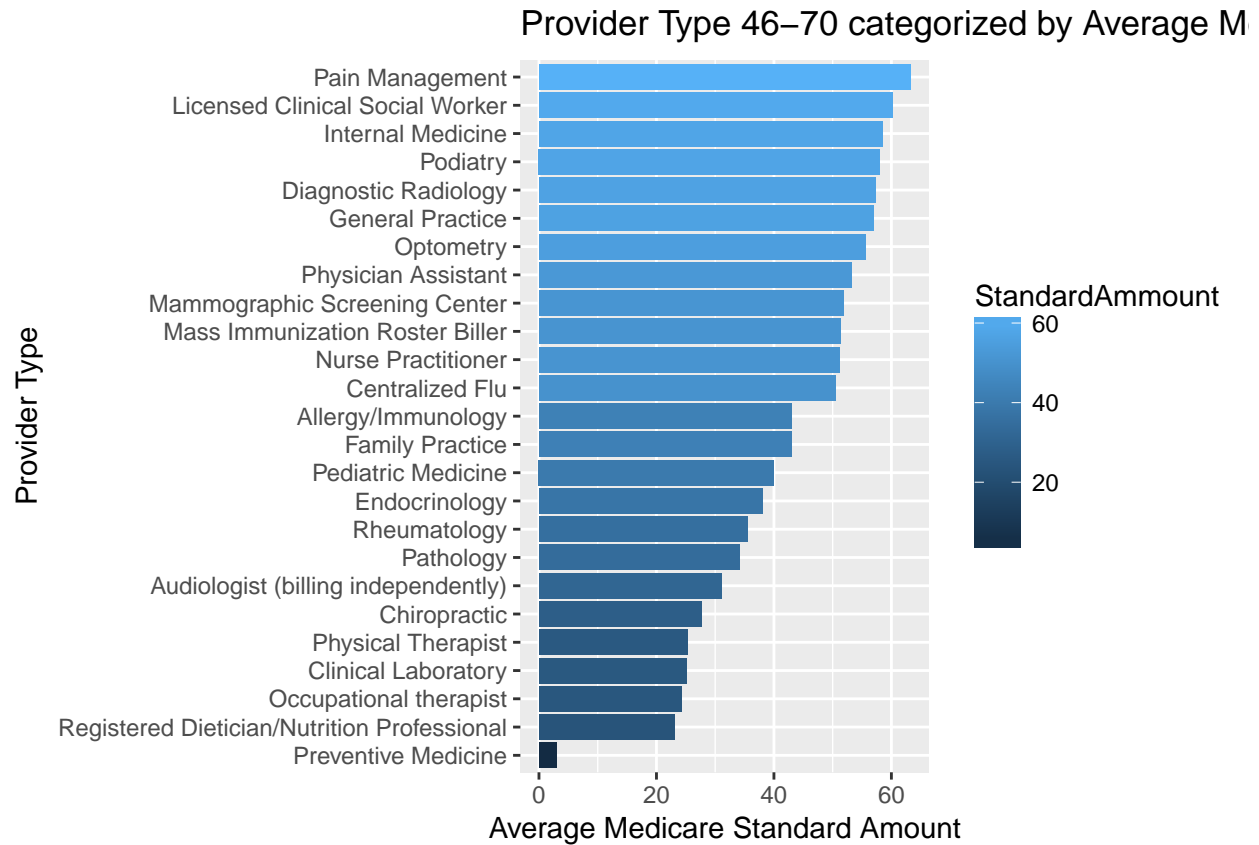


```
ggplot(StandardPs, aes(reorder(Provider, StandardAmount), StandardAmount, fill=StandardAmount)) + geom_bar
```

Provider Type 21–45 categorized by Average Medicare



```
ggplot(StandardPt, aes(reorder(Provider, StandardAmount), StandardAmount, fill=StandardAmount)) + geom_bar
```

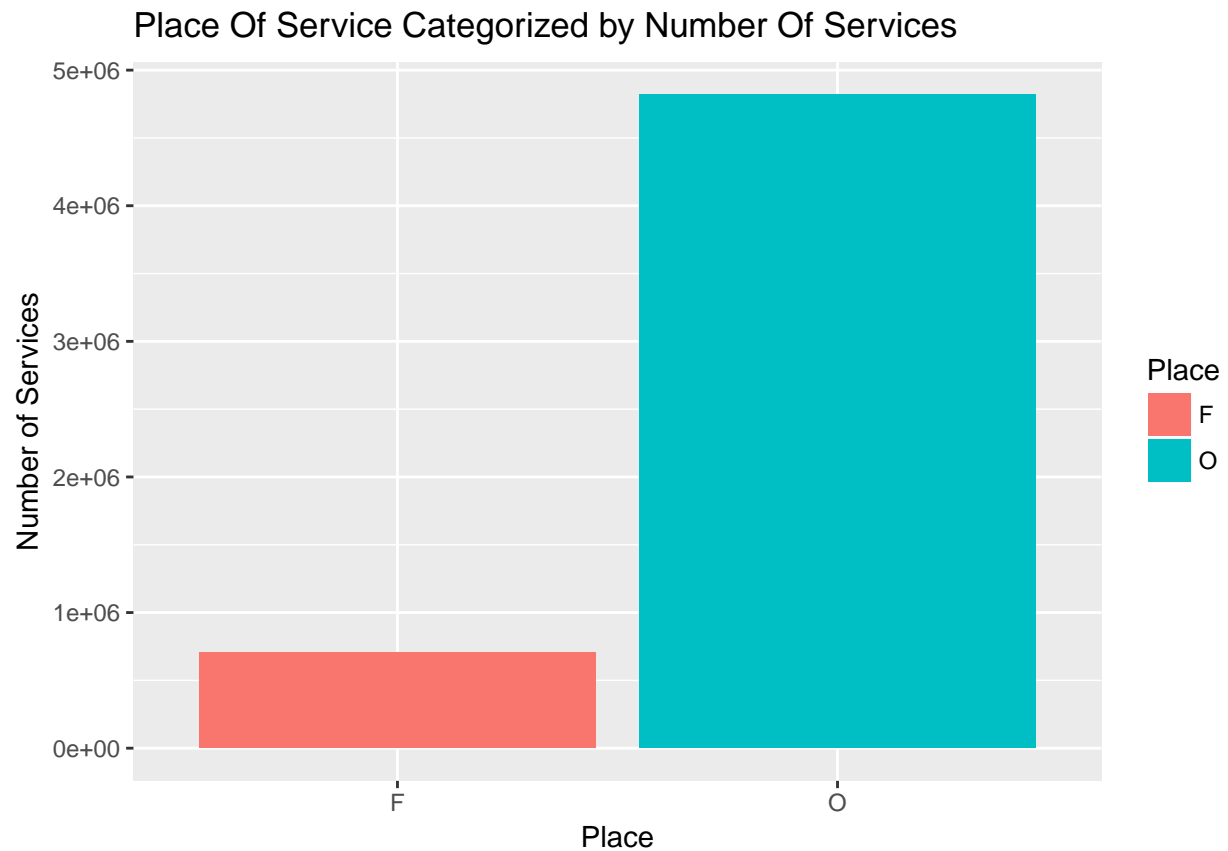



Place of Service

In these graphs, “O” is in reference to providers located in offices, and “F” refers to providers located in facilities.

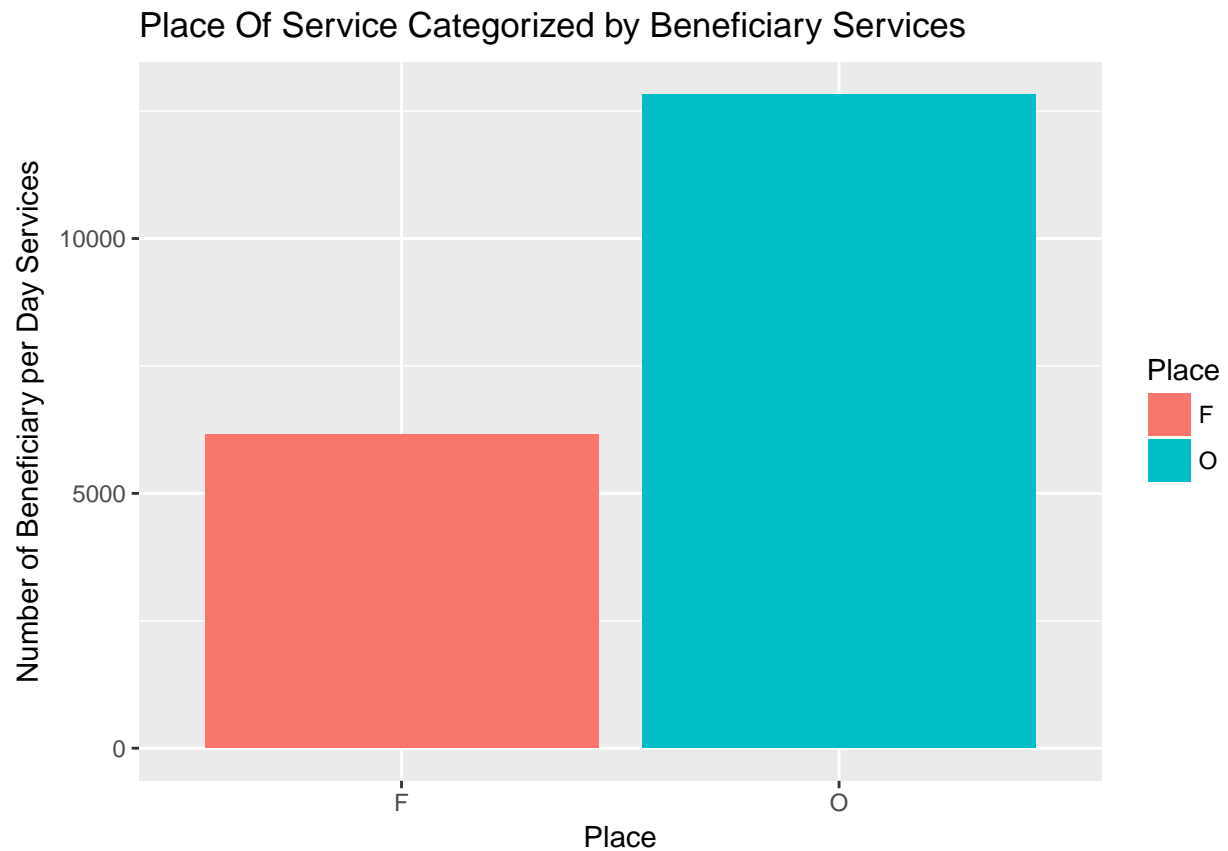
Number of Services and Place of Service

```
PlaceServices=providerspokane%>%group_by(Place=Place.of.Service)%>%summarize(NumberOfServices =sum(NumberOfServices))
ggplot(PlaceServices,aes(Place,NumberOfServices,fill=Place))+geom_bar(stat="identity")+ylab("Number of Services")
```



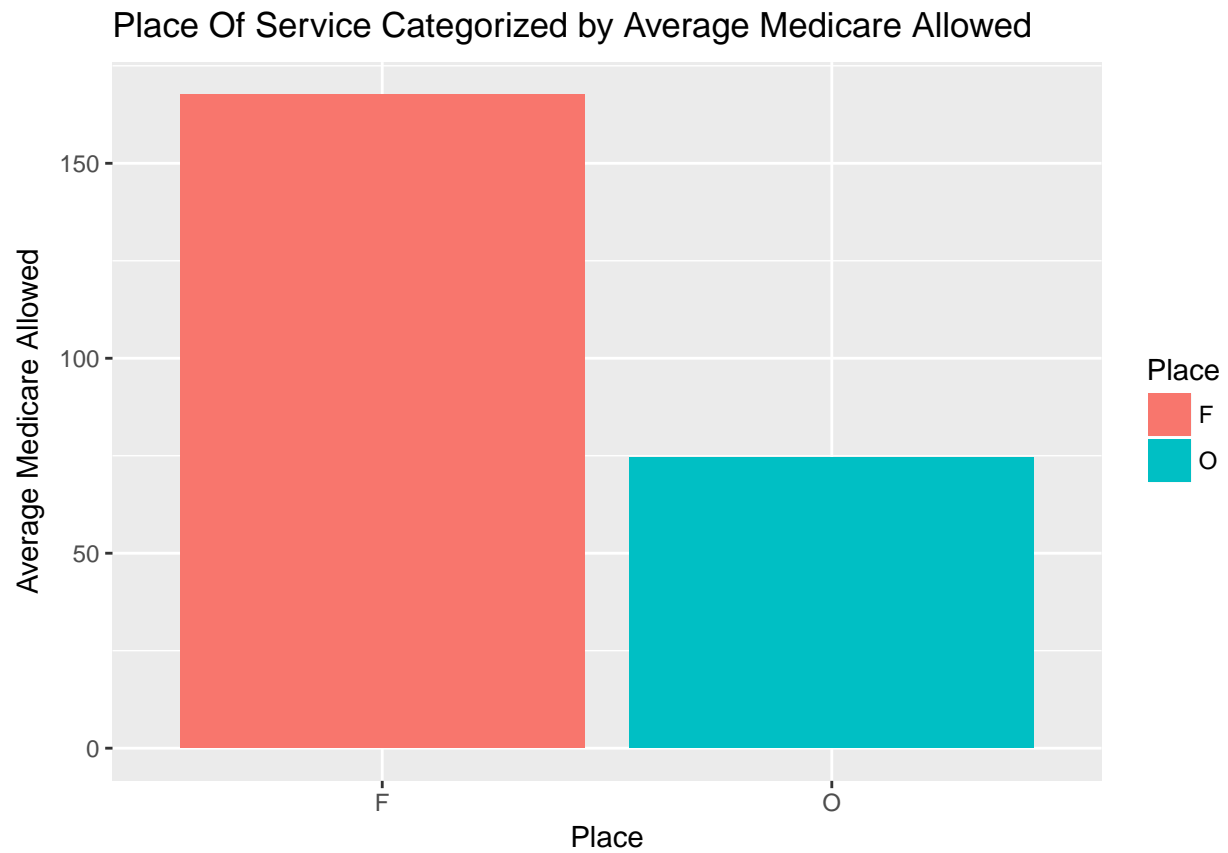
Number of Beneficiary per day Services and Place of Service

```
PlaceBeneficiary=providerspokane%>%group_by(Place =Place.of.Service)%>%summarize(NumberofBeneficiary=length(Beneficiary))
ggplot(PlaceBeneficiary,aes(Place,NumberofBeneficiary,fill=Place))+geom_bar(stat="identity")+ylab("Number of Beneficiary per day Services and Place of Service")
```



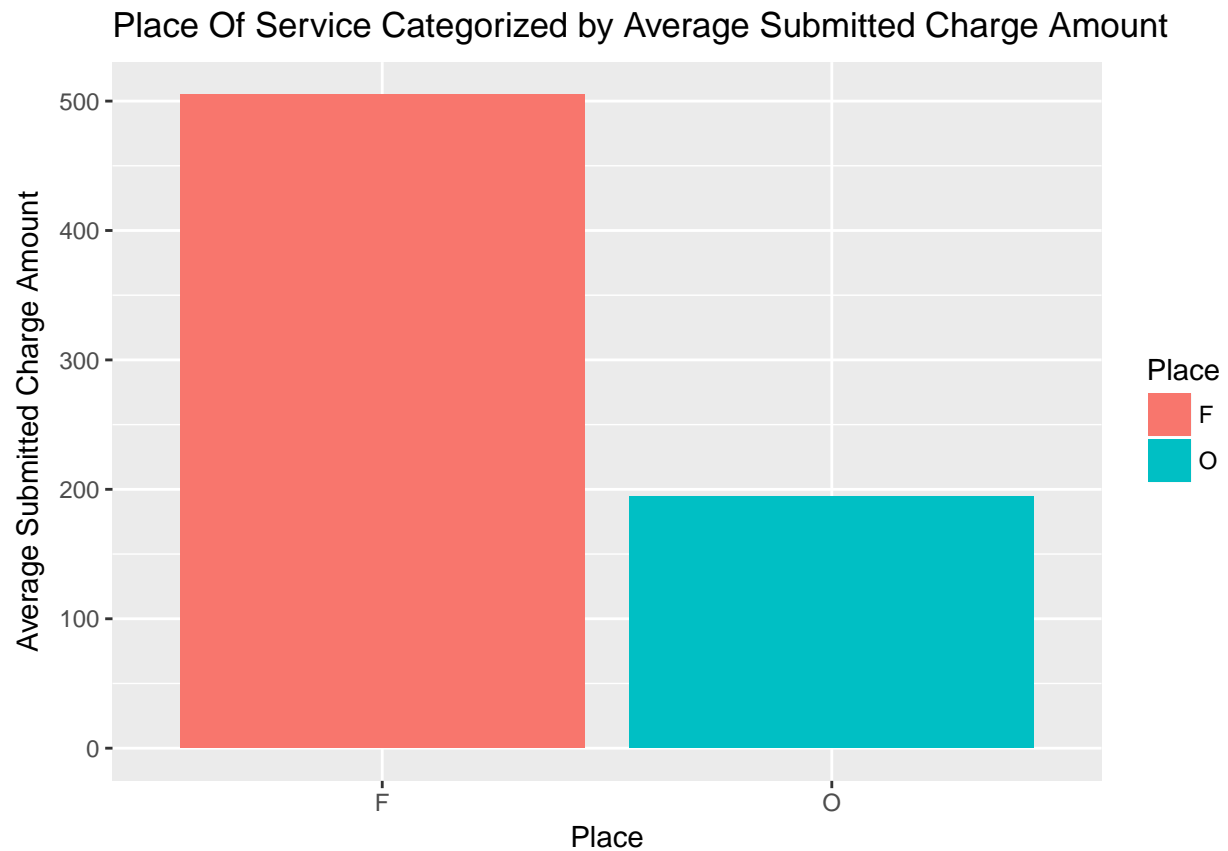
Average Medicare Allowed and Place of Service

```
PlaceAllowed=providerspokane%>%group_by(Place=Place.of.Service)%>%summarize(AllowedAmmount=mean(Average
ggplot(PlaceAllowed,aes(Place,AllowedAmmount,fill=Place))+geom_bar(stat="identity")+ylab("Average Medic
```



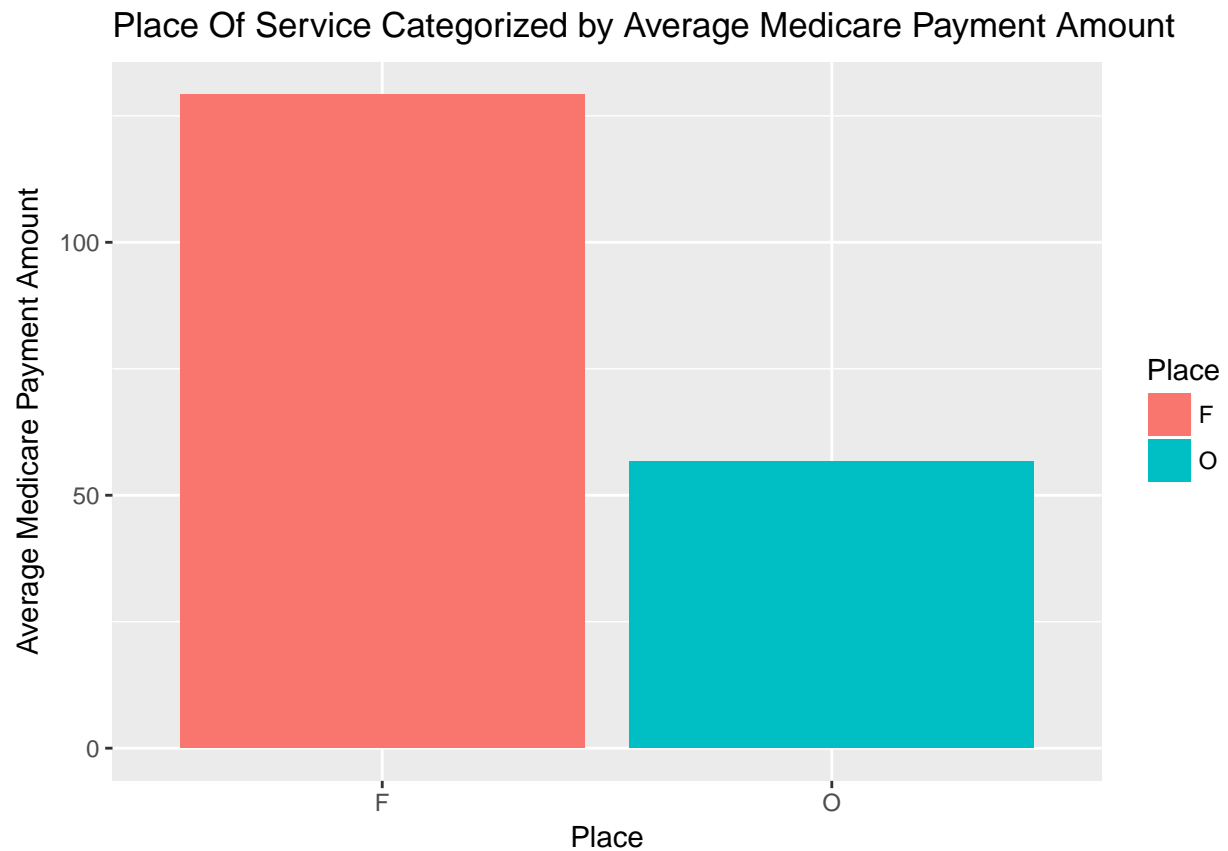
Average Submitted Charge Amount and Place of Service

```
PlaceCharged=providerspokane%>%group_by(Place=Place.of.Service)%>%summarize(ChargedAmmount=mean(Average
ggplot(PlaceCharged,aes(Place,ChargedAmmount,fill=Place))+geom_bar(stat="identity")+ylab("Average Submi
```



Average Medicare Payment Amount and Place of Service

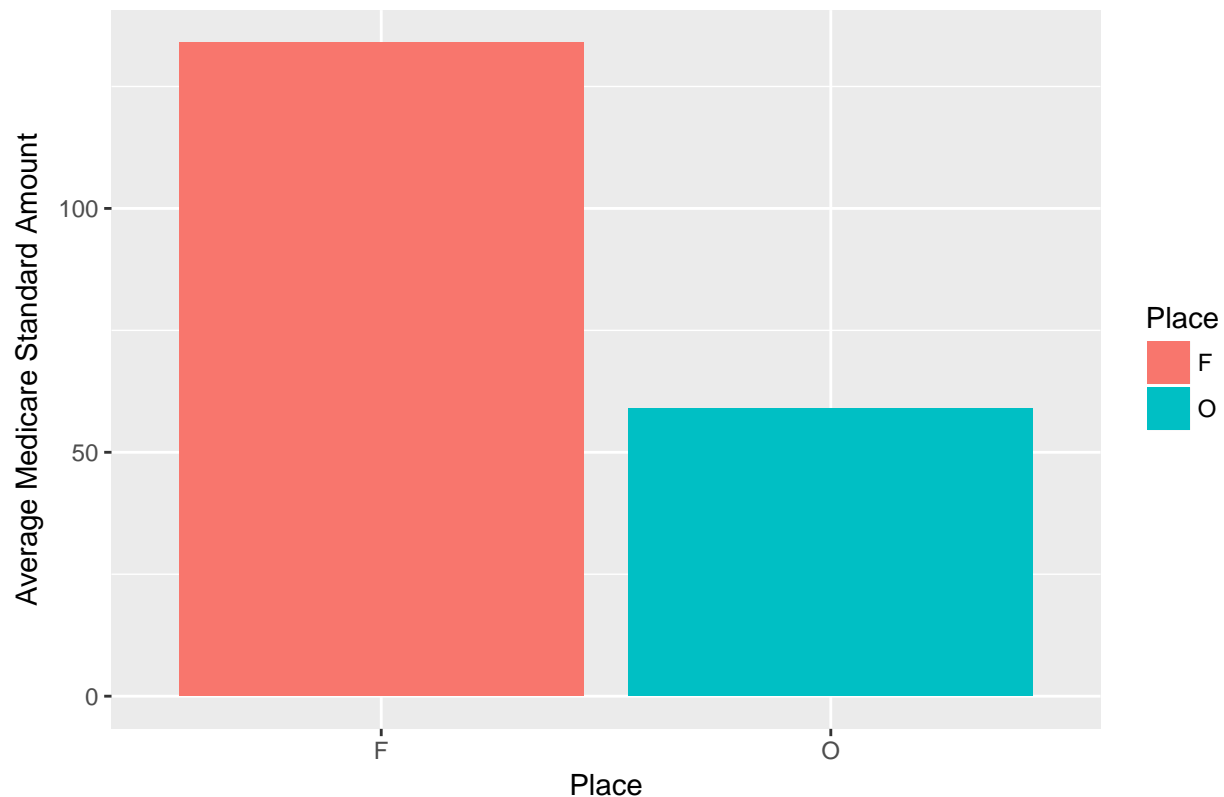
```
PlacePayment=providerspokane%>%group_by(Place=Place.of.Service)%>%summarize(PaymentAmount=mean(Average
ggplot(PlacePayment,aes(Place,PaymentAmount,fill=Place))+geom_bar(stat="identity")+ylab("Average Medicare Payment Amount by Place of Service")
```



Average Medicare Standardized Amount and Place of Service

```
PlaceStandard=providerspokane%>%group_by(Place=Place.of.Service)%>%summarize(StandardAmmount=mean(Avera
ggplot(PlaceStandard,aes(Place,StandardAmmount,fill=Place))+geom_bar(stat="identity")+ylab("Average Med
```

Place Of Service Categorized by Average Medicare Standardized Amount



Question 2

How do all these variables vary as a function of the type of service that is being provided?

For the gender variable, we had three categories: females, males, and non-specified. For the non-specified gender, we made these responses N/A and removed them, as they were skewing the averages of our data with the small sample and large outliers. In Spokane, Female providers have fewer services than male providers in medical fields, which are likely a result of the amount of female medical professionals there are on Average anywhere. The first bar graph shows this disparity clearly, as you can see that females have only half as many providers as males do in the Spokane area. There are also less people visiting providers whom are female, but this could simply be a result of there being less female providers available. Females generally have less available Medicare coverage for their services than male providers so, which would mean that they are involved in fields that are less common, less expensive, and less immediate and necessary (i.e. surgery, oncology, emergency room services). Because of this, Males submit more charges for their services and have higher payment amounts for Medicare.

For provider type compared to number of services, we put the providers in descending order by number of services provided. We then graphed the providers in chunks, so that we could clearly see each provider, as they were very cramped for space all in one graph and not appealing to the eye. The gradient of blue emphasizes which provider has the highest amount on each graph (getting darker as the number decreases). The provider with the most services is Clinical Labs, as this is a very broad service that does a lot of different things, from research to testing to diagnosis of all kinds. They had about 1,250,000 services; The providers with the most services were mostly primary care doctors and oncologists. Pediatrics and geriatrics had the least services because they are more specialized fields. For average Medicare allowed, the ambulatory surgical center was first, because it is a service (emergency surgery) where the situation is immediate, necessary, and very expensive. Therefore, you need more Medicare coverage for this service. Ambulance services are the

highest submitted charge as well as the highest Medicare coverage, this is because not long are ambulances expensive, but they are the most frequently used service on this list. Because preventative medicine is not exactly necessary, very expensive, or common; it is the lowest in submitted charge and Medicare coverage.

For the place variable, there are two options: facilities or offices. Facilities are buildings such as hospitals and offices are buildings such as a primary care doctor or an orthopedic doctor. Offices are used more often for providers with more services because we can assume that patients needing care from these providers were referred there by facilities such as hospitals who don't specialize in that field. Facilities have higher Medicare coverage and submitted charges because they deal with patients who need immediate, expensive, and vital care such as surgery and ambulances.

Conclusion and Limitations

In our findings, we found that all of our data can be divided essentially into two categories. These are providers in Spokane who have a lot of Medicare coverage and those who have less Medicare coverage. Providers who have a lot of coverage are, on average, Male. They usually run their services out of facilities such as hospitals. These Providers provide fewer services and these services are that are expensive, common, and absolutely necessary to their patients. Providers who have less Medicare coverage were generally Female and worked out of offices. These Providers provide more services and these services are less expensive, more specialized, and less common.

Some limitations to our findings included the unidentified gender variable, our knowledge of the ability to split the provider type variable into visuals that were concise and easy to understand due to the amount of providers we were presented with. We wanted to try to use other types of graphs to show our data to its fullest potential, but felt that bar graphs showed our data the most clearly every time. We had some trouble renaming data, such as the variables in Places of Service, so that it is easier for the reader to understand. A big limitation in the end was our lack of understanding of medical providers and the medical field in general, which made it very difficult to analyze our data in a detailed manner.