۱-

A_i : پیشامد اینکه دانش‌آموز iام روی صندلی با شماره مختص خود بنشیند

وقتی نوبت نفر iام است که بنشیند، در همه حالات، صندلی‌های ۲ تا (i -1) پر هستند و همچنین یک صندلی دیگر از بین (62 - i) تا صندلی باقیمانده توسط دقیقاً یکی از دانش‌آموزان با شماره کارت‌های ۱ تا (i - 1) پر می‌شود.که هر کدام ازین دانش‌آموزها شانس یکسانی برای نشستن روی هر کدام از این (62-i) صندلی دارند. پس احتمال نشستن دانش‌آموز iام روی صندلی خود برابر است با:

$$P(A_i) = \frac{1}{62 - i}$$

بنابراین، احتمال اینکه دانش‌آموز آخر(نفر ۶۰م) روی صندلی خود بشیند برابر خواهد بود با:

$$P(A_i) = \frac{1}{62 - 60} = \frac{1}{2}$$

۲-احتمال اینکه همه‌ی کارمندها تولد یکسان داشته باشند یا به بیان دیگر احتمال اینکه همه‌ی کارمندها ۳۶۴ روز از ۳۶۵ روز سال را کار کنند، برابر خواهد بود با: $P(n) = 365 \times (\frac{364}{365})^n$

ضریب ۳۶۵ در فرمول بالا به این دلیل است که این روز تولد n کارمند می‌تواند هر یک از ۳۶۵ روز سال باشد.

حال می‌توان میانگین تعداد روزهای سال که همه‌ی کارمندها در آن کار می‌کنند را با امید ریاضی متغیر n نشان داد:

$$E(N) = \sum_{x=0}^{n} x.p(x) = p(1) + 2p(2) + \cdots + np(n)$$

برای ماکسیمم کردن میزان productivity شرکت، باید n را بگونه‌ای بیابیم که امید ریاضی را بیشینه کند. پس کافیست n را به ازای وقتی که مشتق امید ریاضی مساوی صفر است، محاسبه کنیم:

$$\frac{d\,E(n)}{dn} = 0 \;\rightarrow\; \frac{d\,E(n)}{dn} = \frac{d\left(nP(n)\right)}{dn} = 1 \times P(n) + n \times \frac{d\,P(n)}{dn} = P(n) + n \times \left(365 \times \left(\frac{364}{365}\right)^n\right)' =$$

$$= 365 \times \left(\frac{364}{365}\right)^n + n \times 365 \times \left(\frac{364}{365}\right)^n \times \ln\left(\frac{364}{365}\right)$$

$$= 365 \times \left(\frac{364}{365}\right)^n \left[1 + n \times \ln\left(\frac{364}{365}\right)\right] = 0 \;\rightarrow\; n \times \ln\left(\frac{364}{365}\right) = -1 \;\rightarrow\; n = 364.5$$

پس تعداد کارمندان باید ۳۶۴ یا ۳۶۵ باشد.

۳-

a) اگر یک آزمایش تصادفی به شکلی باشد که وقوع یک پیشامد، مرتبط با واحد مکان یا زمان باشد، یک آزمایش پواسن تشکیل شده است. یک متغیر تصادفی پواسن یک متغیر تصادفی است که می‌تواند برای شمارش تعداد پیشامدهای گسسته که در فاصله زمانی مشخصی اتفاق می‌افتند بکار رود. به طوریکه رخداد این پیشامدها بین هر بازه زمانی مجزا مستقل از هم باشند. فروش یا عدم فروش هر پیتزا توسط فروشنده به فروش پیتزای قبلی یا بعدی وابسته نیست. همچین داریم فروش پیتزا توسط بندیک را در یک بازه زمانی مشخص(a round of his route) می‌کنیم. پس تعداد پیتزاهای فروخته شده توسط بندیک در بازه زمانی مشخص شده از توزیع پواسن پیروی می‌کند. (پارامتر λ برابر با ۲۰ می‌باشد)

b)

$$P(X\ is\ even) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^{2k}}{(2k)!} = e^{-\lambda}(\frac{\lambda^0}{0!} + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \cdots + \frac{\lambda^{2k}}{(2k)!}) = e^{-\lambda}\cosh(\lambda)$$
$$= e^{-\lambda}\frac{1}{2}(e^{-\lambda} + e^{\lambda}) = \frac{1}{2}(1 + e^{-2\lambda}) = \frac{1 + e^{-40}}{2}$$

c) ابتدا مقدار x را پیدا کردم که به ازای آن توزیع پواسن با پارامتر ۲۰، مقدار تقریبا ۱ را می‌گیرد. سپس تابع توزیع تجمعی احتمال را به ازای مقادیر مثبت از 0 تا آنx یافته شده حساب کردم.

```
#c)calculating the chance of Benedict Selling an even number of pizzas
lim <- qpois(0.9999999999999999,20)
res <- 0
for (i in seq(0,lim, by= 2)) {
  res <- res + dpois(i,20)
}
print(res)
```

The result:

```
> print(res)
[1] 0.5
>
```

۴-

a)متغیرهای تصادفی X_A و X_B از نوع برنولی هستند. چرا که برای توزیع هر کدام از آنها برابر خواهد بود با:

$$P(X_A) = \binom{NP}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{NP-k}$$

$$P(X_B) = \binom{N(1-P)}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{N(1-P)-k}$$

و با استفاده از فرمول های میانگین و واریانس برای متغیرهای تصادفی برنولی داریم:

$$\mu_{X_A} = \frac{1}{2}NP \quad و \quad \sigma_{X_A} = \sqrt{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)NP} = \frac{\sqrt{NP}}{2}$$

$$\mu_{X_B} = \frac{1}{2}N(1-P) \quad و \quad \sigma_{X_A} = \sqrt{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)N(1-P)} = \frac{\sqrt{N(1-P)}}{2}$$

----

b) همانطور که در قسمت (a) محاسبه کردیم، میانگین افرادی که به تاریخ A رای دادند برابر با $\frac{1}{2}NP$ می‌باشد. همچنین میانگین کل افرادی که رای دادند برابر خواهد بود حاصلضرب تعداد کل افراد در احتمال رای دادن هر فرد(0.5). یعنی: $\mu = \frac{1}{2}N$

حال برای محاسبه حد بخشی از دانشجوها که به تاریخ A رای دادند زمانی که N به سمت بی‌نهایت می‌رود داریم:

$$\lim_{N\to+\infty} \frac{\frac{1}{2}NP}{\frac{1}{2}N} = P$$

----

c)توزیع متغیرهای تصادفی X_A و X_B این بار برابر خواهد بود با:

$$P(X_A) = P(X_A{}') + P(X_A{}'') = \binom{q_A NP}{k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{q_A NP-k} + \binom{(1-q_A)NP}{k}\left(\frac{3}{4}\right)^k\left(\frac{1}{4}\right)^{(1-q_A)NP-k}$$

$$P(X_B) = P(X_B{}') + P(X_B{}'') =$$
$$= \binom{q_B N(1-P)}{k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{q_B NP-k} + \binom{(1-q_B)N(1-P)}{k}\left(\frac{3}{4}\right)^k\left(\frac{1}{4}\right)^{(1-q_B)NP-k}$$

آنگاه، برای محاسبه میانگین(امید ریاضی) آنها داریم:

$$E[X_A] = E[X_A{}' + X_A{}''] = E[X_A{}'] + E[X_A{}''] = \frac{1}{4}q_A NP + \frac{3}{4}(1-q_A)NP$$

$$E[X_B] = E[X_B{}' + X_B{}''] = E[X_B{}'] + E[X_B{}''] = \frac{1}{4}q_B N(1-P) + \frac{3}{4}(1-q_B)N(1-P)$$

----

d) میانگین افرادی که به تاریخ A رای دادند در حالت تعمیر سایت را در قسمت (c) محاسبه کردیم. میانگین کل افرادی که رای دادند در حالت تعمیر سایت برابرN خواهد بود چرا که سایت دیگر کسی را بیرون نمی‌اندازد. حال برای محاسبه حد داریم:

$$\lim_{N\to+\infty}\frac{\frac{1}{4}q_A NP+\ \frac{3}{4}(1-q_A)NP}{N}=\frac{1}{4}q_A P+\ \frac{3}{4}(1-q_A)P=P*\left(\frac{1}{4}q_A-\frac{3}{4}q_A+\frac{3}{4}\right)$$
$$=\left(\frac{3}{4}-\frac{1}{2}q_A\right)P$$

پس در این حالت برابر P نمی‌شود.

۵-

a) وقتی دو متغیر تصادفی ناسازگارند(disjoint)، یعنی هرگز امکان ندارد که پیشامدهای مربوط به هر یک از آنها باهم اتفاق بیفتد. در حالیکه وقتی دو متغیر تصادفی مستقل از هم هستند(independent)، دانستن مقدار یکی از آنها هیچ اطلاعاتی در مورد فهمیدن مقدار دیگری در اختیار ما نمی‌گذارد. وقتی دو متغیر تصادفی ناسازگارند، اتفاقا مستقل نیستند. به طور مثال با فرض ناسازگار بودن دو پیشامد A و B، هر کدام از آنها دارند اطلاعاتی راجب دیگری به ما می‌دهند. اینکه اتفاق افتادن B، به این معنی است که A اتفاق نمیفتد(و برعکس)

---

b) بله – اگر احتمال رخ دادن حداقل یکی از این دو پیشامد به ازای همه نقاط فضای نمونه صفر باشد. در این صورت با توجه به ناسازگار بودن دو پیشامد(یعنی احتمال رخ دادن آنها باهم صفر است) و با توجه به توضیح بالا داریم:

$$P(A\cap B)=0=P(A)P(B)$$

---

c) برای اینکه پیشامدهای A و B مستقل از هم باشند، کافیست داشته باشیم:

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=P(A)\ \rightarrow\ P(A\cap B)=P(A)P(B)\ \ (۱)$$

حال هر کدام از احتمال ها را محاسبه می‌کنیم:

$$P(A\cap B)=\frac{\binom{n}{1}}{2^n}=\frac{n}{2^n}$$
$$P(A)=\frac{\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n-1}}{2^n}=\frac{2^n-\binom{n}{0}-\binom{n}{n}}{2^n}=\frac{2^n-2}{2^n}=1-\frac{1}{2^{n-1}}$$
$$P(B)=\frac{\binom{n}{0}+\binom{n}{1}}{2^n}=\frac{n+1}{2^n}$$

حال در تساوی (۱) جایگذاری می‌کنیم:

$$\frac{n}{2^n}=\left(1-\frac{1}{2^{n-1}}\right)\frac{n+1}{2^n}\ \rightarrow\ \frac{n}{n+1}=1-\frac{1}{2^{n-1}}\ \rightarrow\ \frac{1}{2^{n-1}}=1-\frac{n}{n+1}=\frac{1}{n+1}\ \rightarrow\ 2^{n-1}=n+1\ \ (۲)$$

با قرار دادن مقادیر مختلف به جای n در تساوی (۲) ، مقدار ۳ برای n بدست می‌آید. ($n=3$)

۶-

A : پیشامد پسر بودن حداقل یکی از دو فرزند خانواده

B : پیشامد اینکه حداقل یکی از فرزندان پسری به نام <علی> باشد

M : پیشامد اینکه هر دو فرزند خانواده پسر باشند

N : پیشامد اینکه نام فرزند پسر <علی> باشد

a) در این حالت باید احتمال P(M|A) را محاسبه کنیم:

$$P(M|A) = \frac{P(M \cap A)}{P(A)} = \frac{\binom{2}{2}}{\binom{2}{1} + \binom{2}{2}} = \frac{1}{3}$$

b) در این حالت باید احتمال P(M|B) را محاسبه کنیم. باید دقت داشت که پیشامد B خود اشتراک دو پیشامد A و N می‌باشد.(می‌دانیم P(N) = α) پس ابتدا احتمال P(B) را محاسبه می‌کنیم. و سپس احتمال رخ دادن پیشامد M به شرط وقوع پیشامد B :

$$P(B) = P(A \ \cap N) = \frac{\binom{2}{1}\alpha + \binom{2}{2}\alpha^2 + \ 2\binom{2}{2}\alpha(1-\alpha)}{2^2} = \frac{2\alpha + \ \alpha^2 + 2\alpha - 2\alpha^2}{4} = \frac{4\alpha - \alpha^2}{4}$$

$$P(M|B) = \frac{P(M \cap B)}{P(B)} = \frac{\frac{\binom{2}{2}\alpha^2 + \ 2\binom{2}{2}\alpha(1-\alpha)}{2^2}}{\frac{4\alpha - \alpha^2}{4}} = \frac{2\alpha - \alpha^2}{4\alpha - \alpha^2} = \frac{2-\alpha}{4-\alpha} = 1 - \frac{2}{4-\alpha}$$

c) بله جواب‌ها متفاوت هستند –چون در قسمت (b) شرط <علی> بودن نام حداقل یکی از پسرها مطرح است. همچنین طبق توضیحات و فرمول‌های نوشته شده در قسمت (a) و (b) می‌توان توجیه این قضیه را دید.

-۷

ابتدا قسمت (b) سوال که حالت کلی این سوال است را توضیح میدم و سپس با استفاده از فرمول های بدست آمده در قسمت (b) ، قسمت (a) را پاسخ می‌دهم.

a) با توجه به توضیحات بیان شده در قسمت (b) داریم (تعداد بازیکنان در اینجا ۱۶تاست. پس n = 4):

احتمال روبه‌رو شدن Federer و Nadal در هر یک از match ها برابر است با: $\frac{1}{8\times15} = \frac{1}{120}$

و به طور کلی احتمال روبه‌رو شدن Federer و Nadal در طول تورنمنت در یکی از matchها برابر است با: $\frac{1}{8}$

---

b) در یک تورنمنت با تعداد 2^n شرکت‌کننده، تعداد جفت تنیسورهای متمایز برابر خواهد بود با: $\frac{1}{2}2^n(2^n-1)$

که Federer و Nadal یکی ازین جفت‌ها می‌باشند.

پس احتمال روبه‌رو شدن هر یک ازین جفت تنیسور ها(در اینجا مثلاً Federer و Nadal) در هر یک از matchها برابر است با:

$$\frac{1}{\frac{1}{2}2^{n}(2^{n}-1)} = \frac{1}{2^{n-1}(2^{n}-1)}$$

از آنجایی که تعداد کل matchهای برگزار شده در تورنمنتی با تعداد 2^n شرکت‌کننده برابر $2^n - 1$ می‌باشد (چون برای حذف هر بازیکن به یک match نیاز است و همه بازیکن‌ها به جز یک نفر حذف می‌شوند) ، احتمال روبه‌رو شدن یک زوج تنیسور مشخص (در اینجا مثلاً Federer و Nadal) برابر خواهد بود با:

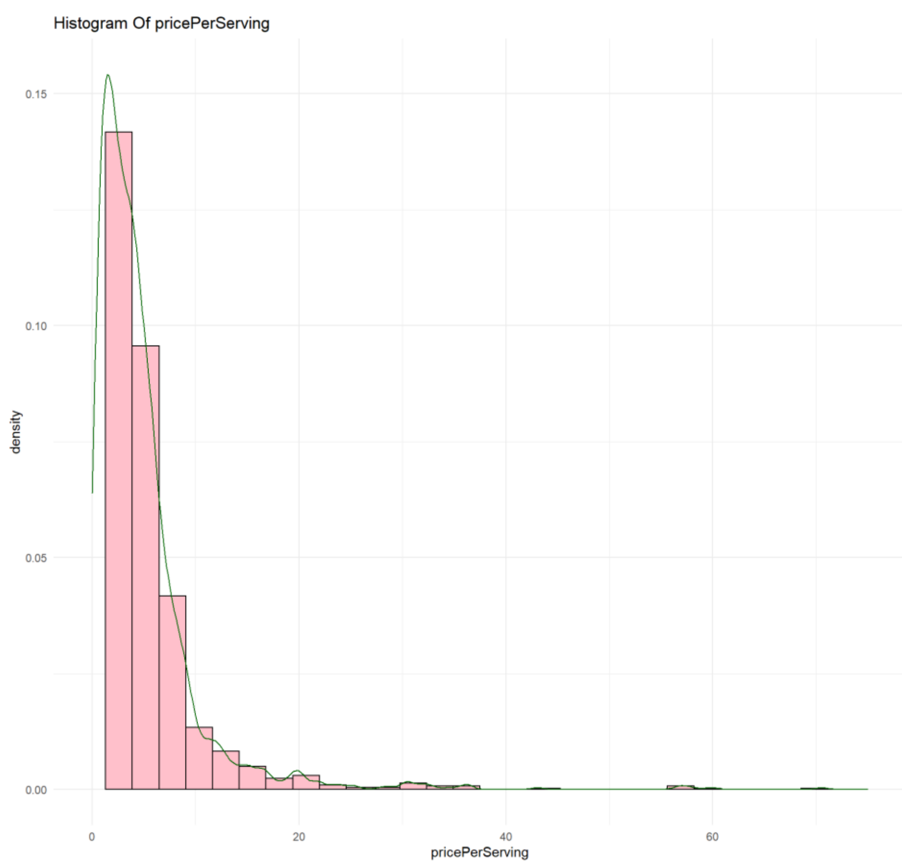
$$\frac{2^{n-1}}{2^{n-1}(2^{n}-1)} = \frac{1}{2^{n-1}}$$

-۸

(a

```r
#a)Drawing the Histogram of pricePerServing with 30 bins(recommended by R)
ggplot(Foods, aes(x=pricePerServing, y = ..density..)) +
  geom_histogram(bins= 30, fill= "pink", col="black")+
  geom_density(col ="darkGreen")+ xlim(0,75)+
  ggtitle("Histogram Of pricePerServing")+ theme_minimal()
```

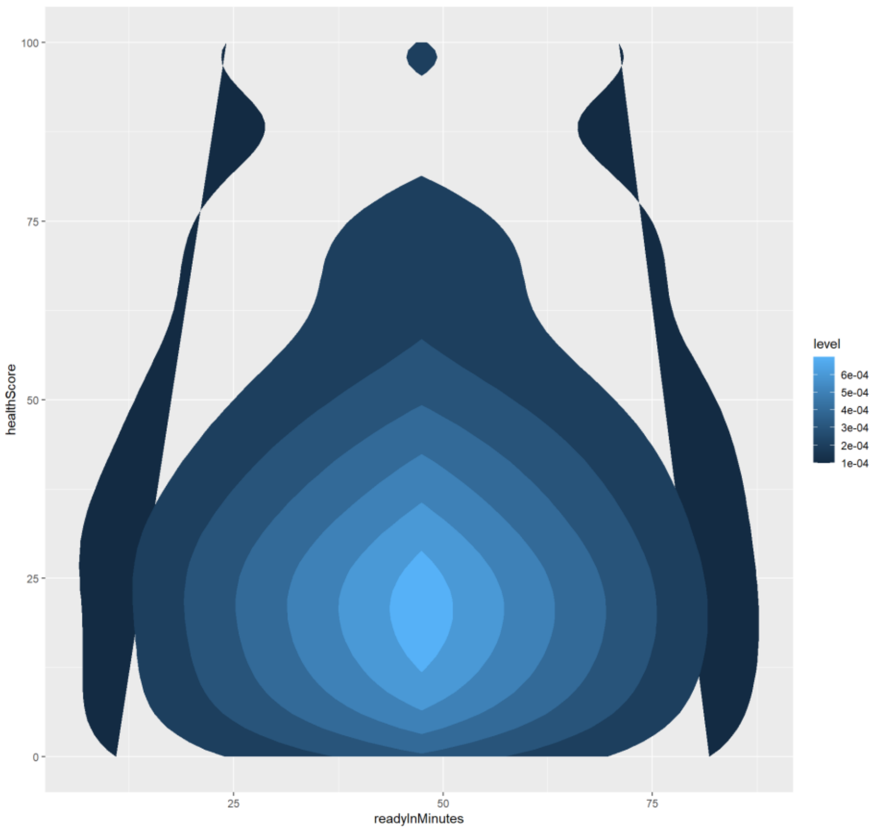
The result:



(b

```
#b)Drawing the 2D density plot
ggplot(Foods, aes(x = healthScore, y = readyInMinutes, fill = ..level..)) +
  stat_density_2d(geom = "polygon")
```

The result:

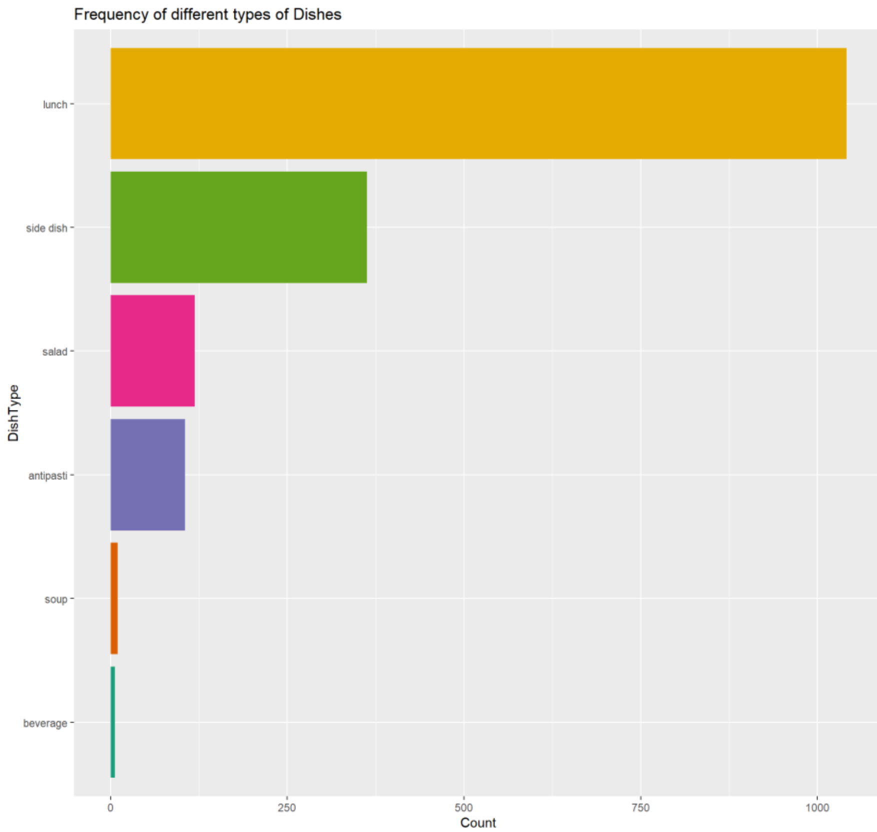


(c

```
#c)Sorting the categories in the dishType and Drawing a horizontal bar plot
sorted <- data.frame(sort(table(Foods$dishType),increasing = TRUE))
colnames(sorted) <- c("DishType","Count")

ggplot(sorted, aes(x= DishType, y=Count , fill=DishType)) +
  geom_bar(stat = "identity" ) +
  scale_fill_brewer(palette="Dark2")+ coord_flip()+
  theme(legend.position = "none")+
  ggtitle("Frequency of different types of Dishes")
```

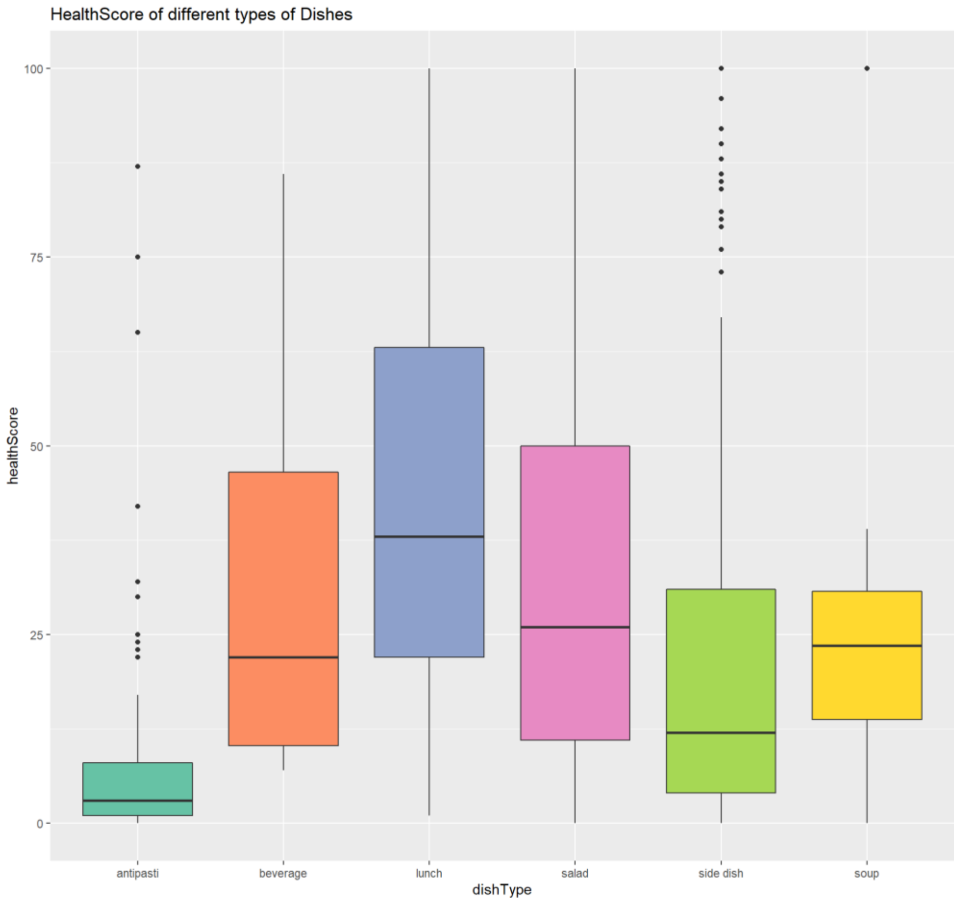
The result:



(d

```
#d)side-by-side boxplots for different dish types
ggplot(Foods, aes(x = dishType, y = healthScore , fill = dishType)) +
  geom_boxplot()+scale_fill_brewer(palette="Set2")+
  theme(legend.position = "none")+
  ggtitle("HealthScore of different types of Dishes")
```

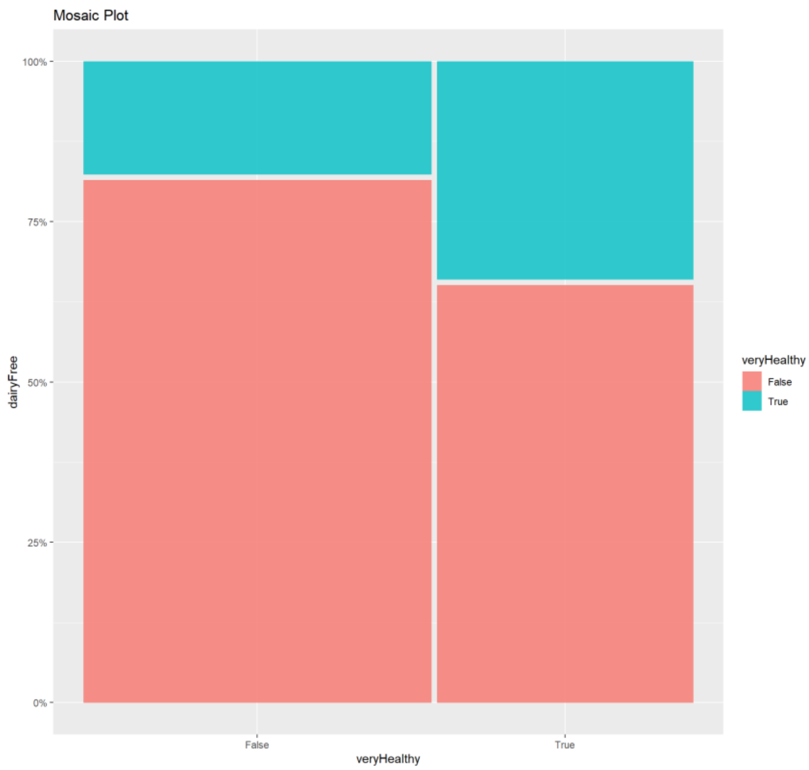
The result:



(e

```
#e)Drawing mosaic plots
ggplot(data = Foods) +
  geom_mosaic(aes(x = product(veryHealthy, dairyFree), fill = veryHealthy)) +
  labs(y="dairyFree", x="veryHealthy", title = "Mosaic Plot") +
  scale_y_continuous(labels=scales::percent)
```

The result:


Mosaic Plot
100%
75%
50%
25%
0%
dairyFree
False
True
veryHealthy
veryHealthy
False
True