고령화연구패널조사(KLoSA) 데이터분석

목 차

1. 개요

- 고령화연구패널조사는 향후 초고령사회로 변화해 가는 과정에서 효과적인 사회경제정책을 수립과 시행에 활용될 기초자료 생산
- 고령사회의 다양한 측면들을 여러 분야의 학제적 연구가 가능한 자료를 구축 및 선진국과 국제비교가 가능한 자료를 생산하고 정책개발과 학술연구의 기초자료로 활용
- 패널조사는 2006년 당시 제주도를 제외한 지역에 거주하는 45세 이상 중고령자 중 일반 가구 거주지를 대상으로 표집 및 조사실시
- 2006년부터 짝수 연도에는 동일한 조사 항목을 중심으로 기본조사를 실시하고, 2007년부터 홀수 연도에는 기본조사에 포함되지 않은 내용을 중심으로 특정 주제를 정하여 조사를 실시
- 조사 대상
  고령화연구패널조사의 대상자는 대한민국 제주도를 제외한 지역에 거주하는 45세(1962년 이전생) 이상의 중고령자 개인을 대상으로
  표본수 약 10,000명을 목표로 표집하여 10,254명이 패널로 구축하고 2014년에는 1962~1963년생을 중심으로 한 920명을 추가

1.1 데이터

Loading

In [1]:

```
library(tidyverse)
library(sqldf)
library(gridExtra)
library(fs)
dir_info("../input")
```

```
── Attaching packages ─────────────────────────────────────── tidyverse 1.2.1 ──
✔ ggplot2 3.1.0.9000     ✔ purrr   0.3.1
✔ tibble  2.0.1          ✔ dplyr   0.8.0.1
✔ tidyr   0.8.3          ✔ stringr 1.4.0
✔ readr   1.3.1          ✔ forcats 0.4.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
```

```
── Conflicts ──────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
Loading required package: gsubfn
Loading required package: proto
Warning message:
"no DISPLAY variable so Tk is not available"Loading required package: RSQLite

Attaching package: 'gridExtra'

The following object is masked from 'package:dplyr':

    combine
```

path	type	size	permissions	modification_time	user	group	device_id	hard_links	special_device_id	inode	block
../input/klosa.csv	file	12.73M	rw-r--r--	2019-04-24 06:38:36	root	root	39	1	0	734602267	104
../input/klosa.rda	file	2.12M	rw-r--r--	2019-04-24 06:38:36	root	root	39	1	0	734602268	104

In [2]:

```
read.csv("../input/klosa.csv") -> klosa
```

1.2 패널

1)패널 참여인원

In [3]:

```
klosa %>% count(year)
```

year	n
2006	10254
2008	8688
2010	7920
2012	7486
2014	7029
2016	6618

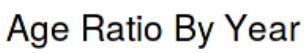
2)패널의 연도별 연령대 비율

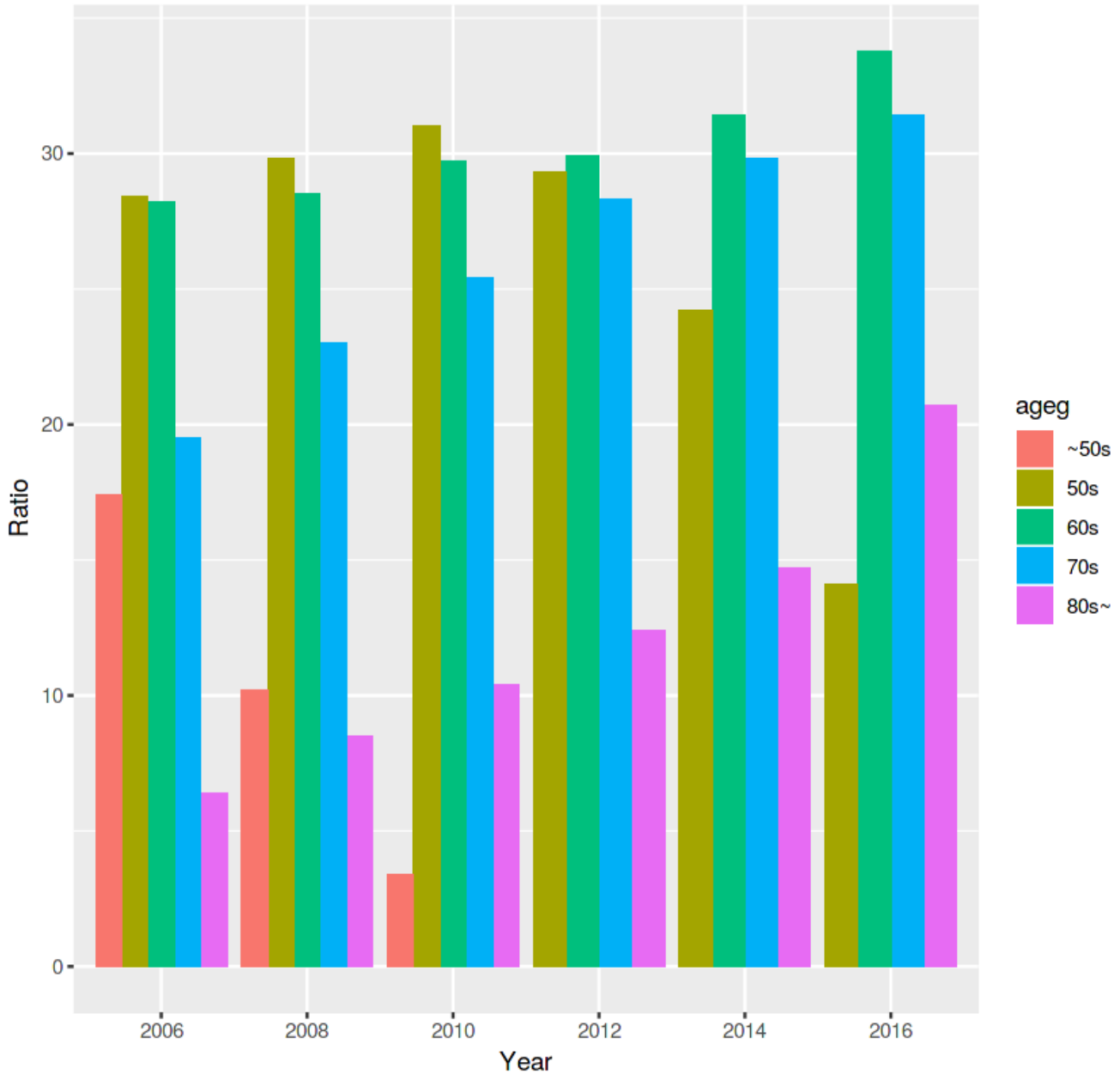
In [4]:

```
klosa %>% filter(!is.na(year)&!is.na(ageg)) %>%
  select(year,ageg) %>%
  group_by(year,ageg) %>%
  summarise(n=n()) %>%
  mutate(total=sum(n)) %>%
  mutate(pct=round(n/total*100,1))->ageg_by_year
```

In [5]:

```
ggplot(ageg_by_year,aes(factor(year) ,pct,fill=ageg))+
  geom_col(position="dodge")+
  ggtitle("Age Ratio By Year")+
  labs(x="Year", y="Ratio")
```





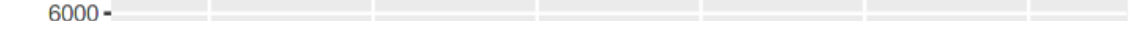
3)패널의 연도별 성별 인원

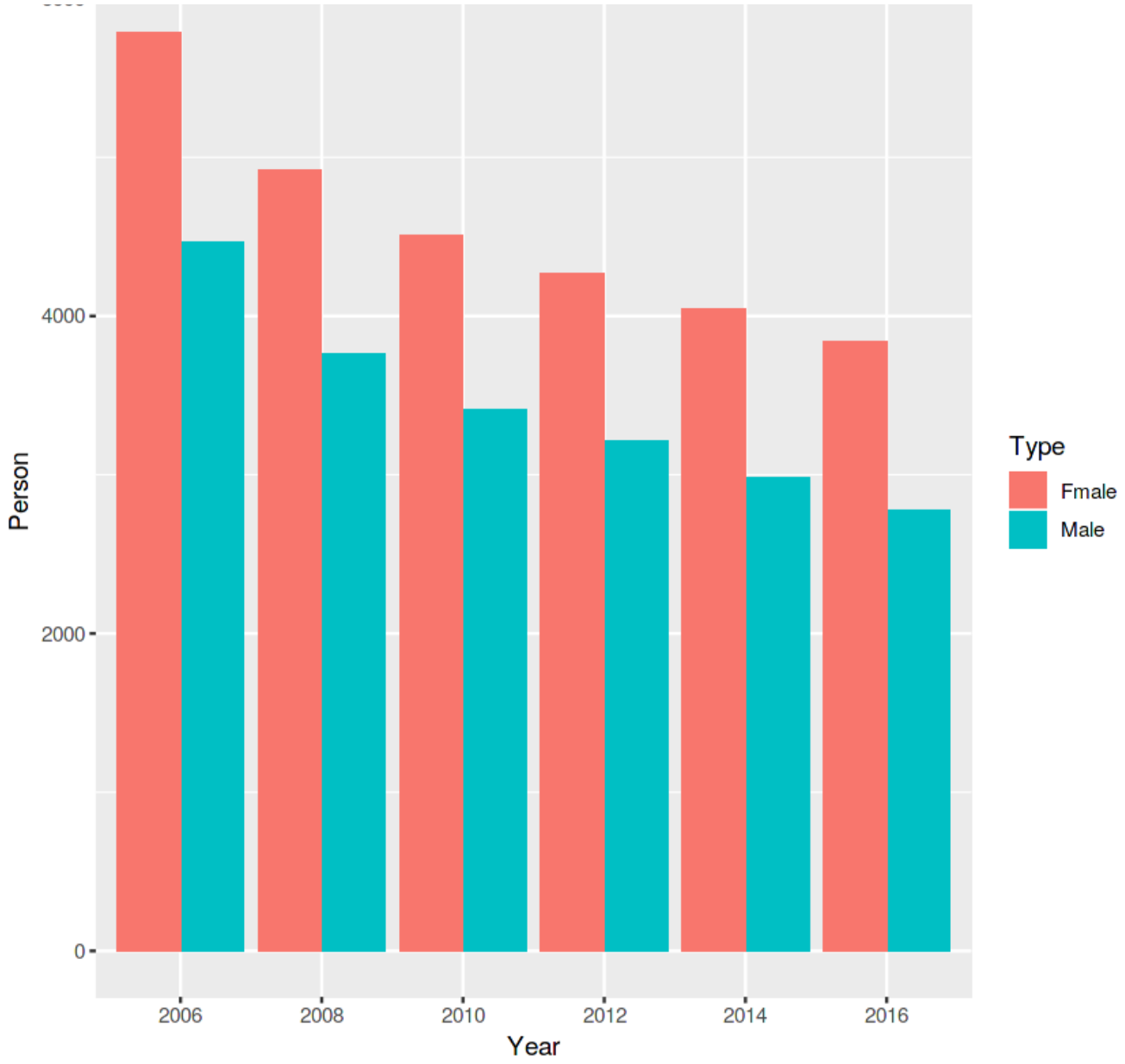
In [6]:

```
klosa %>% filter(!is.na(year)&!is.na(gender)) %>%
  select(year,gender) %>%
  mutate(year=factor(year)) %>%
  group_by(year,gender) %>%
  summarise(n=n()) %>%
  mutate(total=sum(n)) %>%
  arrange(year)-> gender_by_year
```

In [7]:

```
gender_by_year %>% ggplot(aes(year, n, fill=gender))+
  #geom_line()
  geom_col(position='dodge')+
  labs(y="Person",x="Year",title="") +
  theme_grey() +
  scale_fill_discrete(name="Type",
                      labels=c("Fmale","Male"))
```





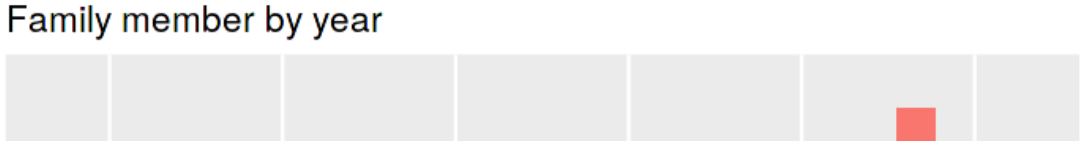
4)패널의 연도별 세대구성 (1세대:1G, 2세대:2G, 3세대:3G, 4세대:4G, 기타:Etc)

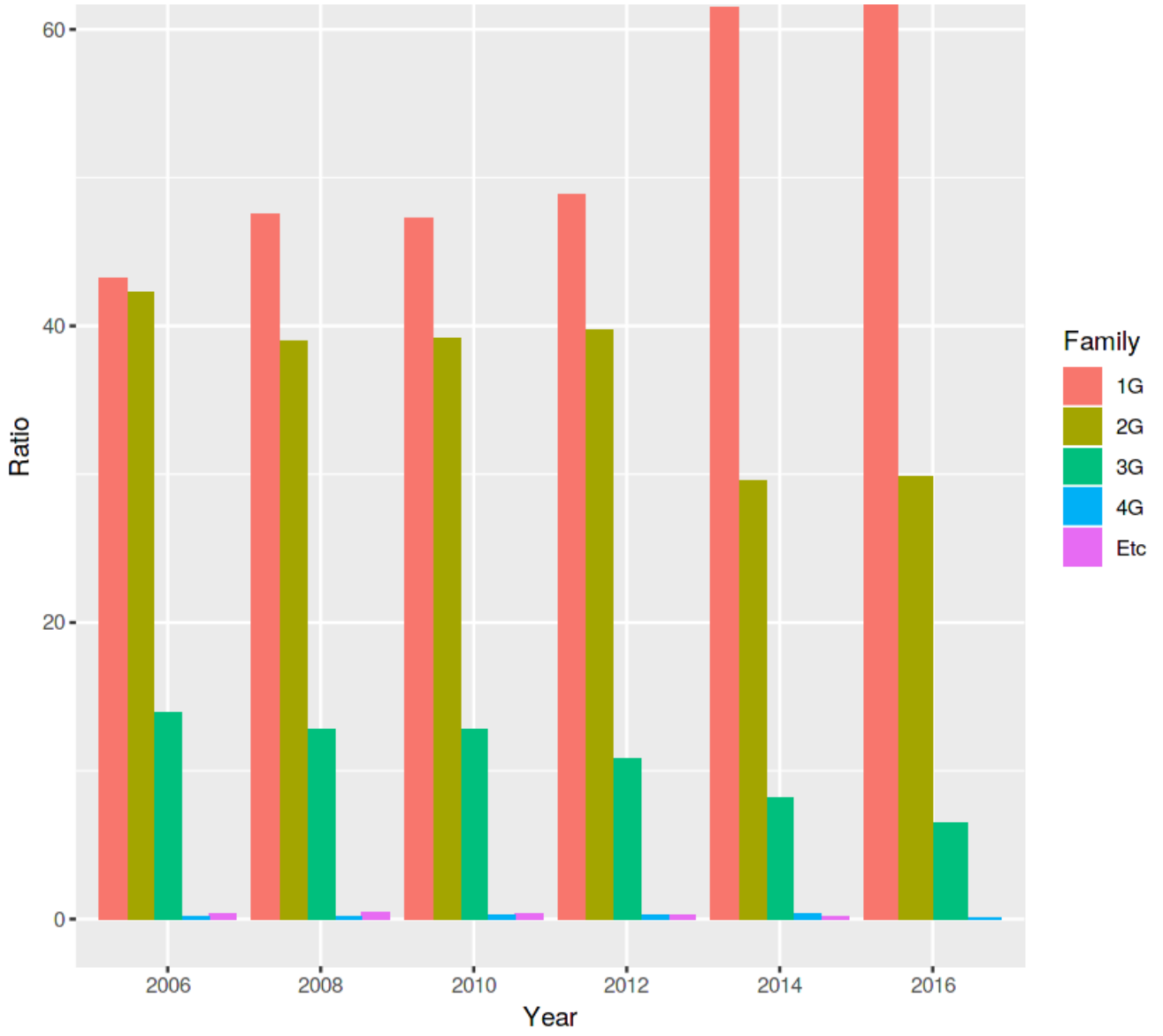
In [8]:

```
klosa %>% filter(!is.na(family)) %>%
  select(year,family) %>%
  group_by(year,family) %>%
  summarise(n=n()) %>%
  mutate(total=sum(n))%>%
  mutate(pct=round(n/total*100,1)) -> family_by_year
```

In [9]:

```
ggplot(family_by_year,aes(factor(year) ,pct,fill=factor(family)))+
  geom_col(position="dodge")+
  ggtitle("Family member by year")+
  labs(x="Year", y="Ratio") +
  theme_grey() +
  scale_fill_discrete(name="Family",
                      labels=c("1G","2G","3G","4G","Etc"))
```





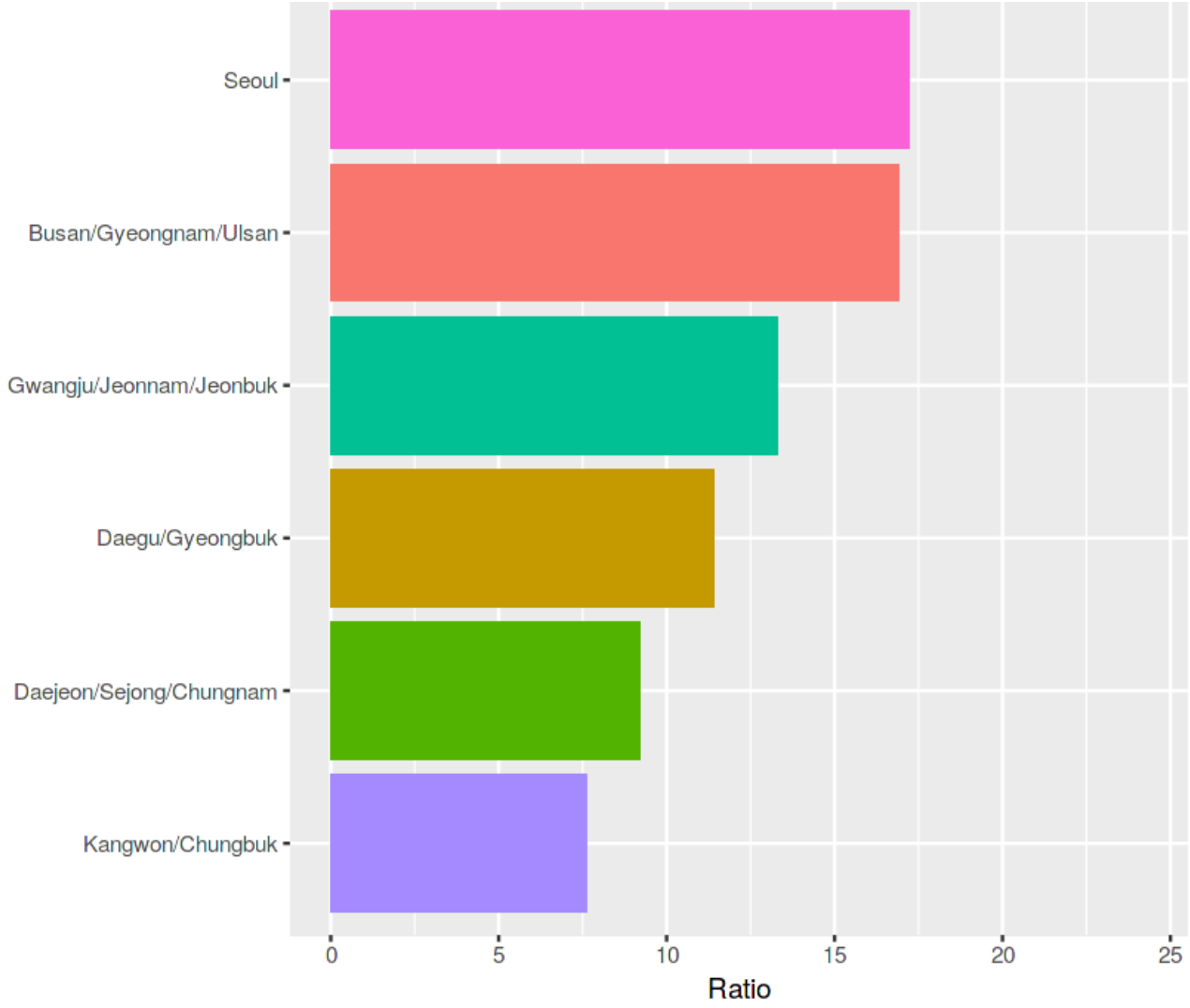
5)패널의 지역구성

In [10]:

```
klosa %>% filter(!is.na(year)&!is.na(regiong)) %>%
  select(year,region,regiong) %>%
  group_by(year,regiong) %>%
  summarise(n=n()) %>%
  mutate(total=sum(n)) %>%
  mutate(pct=round(n/total*100,1)) ->region_by_year
```

In [11]:

```
ggplot(region_by_year[region_by_year$year == "2006",],aes(reorder(regiong,n),pct,fill=regiong))+
  geom_col(position="dodge")+
  coord_flip()+
  ggtitle("Ratio By Region")+
  labs(x="", y="Ratio")+
  guides(fill="none")
```



2. 경제활동

- Employment: 취업
- Jobless: 실업
- Inactive: 비경제제활동

2.1 연도별 연령대별 경제활동

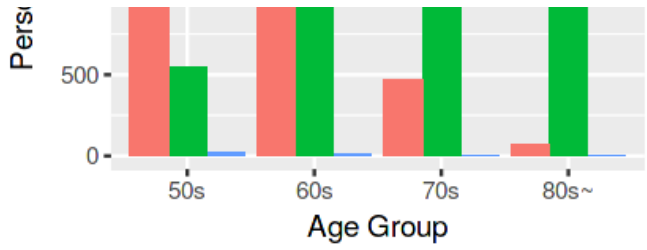
In [12]:

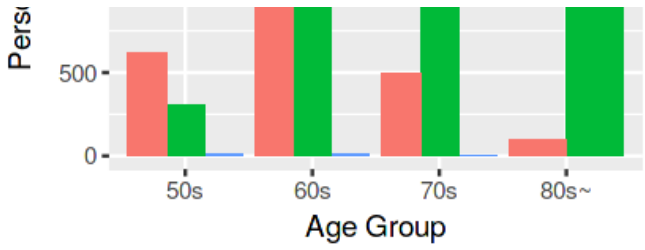
```
klosa %>%filter(!is.na(present_ecotype)) %>%
  select(year,ageg,present_ecotype) %>%
  mutate(year=factor(year)) %>%
  mutate(present_ecotype=ifelse(present_ecotype==1, "Employment",
                                ifelse(present_ecotype==2,"Jobless","Inactive"))) %>%
  count(year,ageg,present_ecotype) %>%
  group_by(year,ageg,present_ecotype) ->ecoact_ageg_year
```

In [13]:

```
eap1<- ggplot(ecoact_ageg_year[ecoact_ageg_year$year == "2006",], aes(ageg, n, fill=present_ecotype
)) +
       geom_col(position='dodge')+
       labs(y="Person",x="Age Group",title="2006 Economic Activity") +
       theme_grey()+
       theme(legend.position = "bottom")+
       scale_fill_discrete(name = "Type")

eap2<- ggplot(ecoact_ageg_year[ecoact_ageg_year$year == "2008",], aes(ageg, n, fill=present_ecotype
```

```
)) +
       geom_col(position='dodge')+
       labs(y="Person",x="Age Group",title="2008 Economic Activity") +
       theme_grey()+
       guides(fill="none")


eap3<- ggplot(ecoact_ageg_year[ecoact_ageg_year$year == "2010",], aes(ageg, n, fill=present_ecotype
)) +
       geom_col(position='dodge')+
       labs(y="Person",x="Age Group",title="2010 Economic Activity") +
       theme_grey() +
       guides(fill="none")

eap4<-ggplot(ecoact_ageg_year[ecoact_ageg_year$year == "2012",], aes(ageg, n, fill=present_ecotype)
) +
      geom_col(position='dodge')+
      labs(y="Person",x="Age Group",title="2012 Economic Activity") +
      theme_grey() +
      guides(fill="none")


eap5<- ggplot(ecoact_ageg_year[ecoact_ageg_year$year == "2014",], aes(ageg, n, fill=present_ecotype
)) +
       geom_col(position='dodge')+
       labs(y="Person",x="Age Group",title="2014 Economic Activity") +
       theme_grey() +
       guides(fill="none")

eap6<- ggplot(ecoact_ageg_year[ecoact_ageg_year$year == "2016",], aes(ageg, n, fill=present_ecotype
)) +
       geom_col(position='dodge')+
       labs(y="Person",x="Age Group",title="2016 Economic Activity") +
       theme_grey()  +
       guides(fill="none")
grid.arrange(eap1, eap2, eap3, eap4, eap5, eap6, ncol=2, nrow = 3)
```





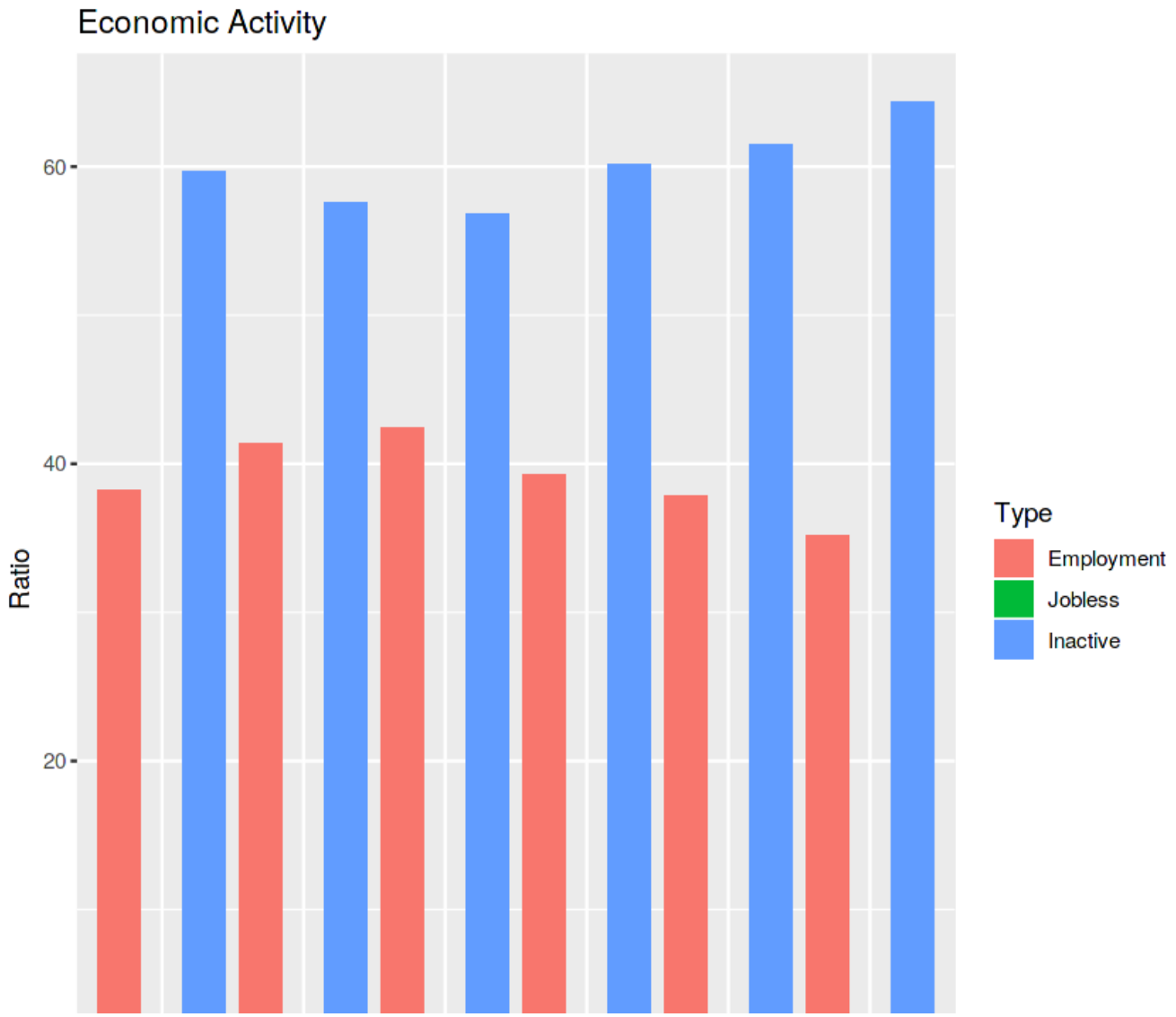
2.2 연도별 경제활동상태

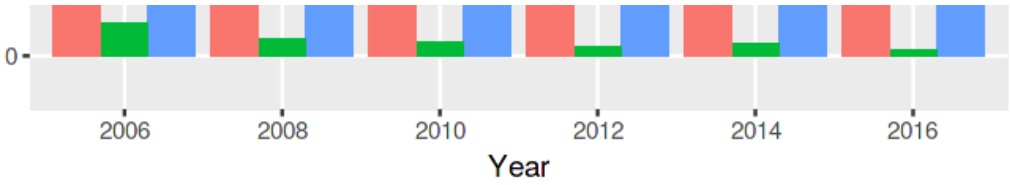
In [14]:

```
klosa %>% filter(!is.na(present_ecotype)) %>%
    mutate(year=factor(year)) %>%
    select(year,present_ecotype) %>%
    count(year,present_ecotype) %>%
    group_by(year) %>%
    mutate(total=sum(n)) %>%
    mutate(pct=round(n/total*100,1))->ecoact_year
```

In [15]:

```
ggplot(ecoact_year,aes(year, pct,fill=as.factor(present_ecotype))) +
  geom_col(position='dodge')+
  labs(y="Ratio",x="Year",title="Economic Activity") +
  theme_grey() +
  scale_fill_discrete(name="Type",
                      labels=c("Employment","Jobless","Inactive"))
```





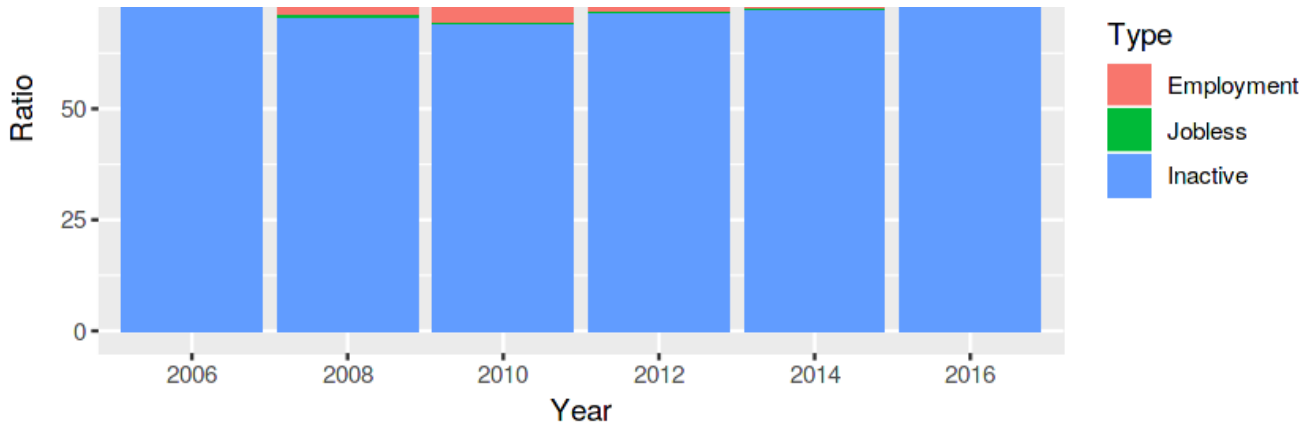
2.3 연도별 성별 경제활동

In [16]:

```
klosa %>% filter(!is.na(gender)&!is.na(present_ecotype)) %>%
  mutate(year=factor(year)) %>%
  select(year,gender,present_ecotype) %>%
  group_by(year,gender,present_ecotype) %>%
  summarise(n=n()) %>%
  mutate(total=sum(n)) %>%
  mutate(pct=round(n/total*100,1)) ->ecoact_gender_year
```

In [17]:

```
eag1<-ggplot(ecoact_gender_year[ecoact_gender_year$gender=="male",],aes(year, pct ,fill=as.factor(p
resent_ecotype))) +
  geom_col()+
  labs(y="Ratio",x="Year",title="Male Economic Activity by year") +
  theme_grey() +
  scale_fill_discrete(name="Type",
                      labels=c("Employment","Jobless","Inactive"))

eag2<-ggplot(ecoact_gender_year[ecoact_gender_year$gender=="female",],aes(year, pct
,fill=as.factor(present_ecotype))) +
  geom_col()+
  labs(y="Ratio",x="Year",title="Female Economic Activity by year") +
  theme_grey() +
  scale_fill_discrete(name="Type",
                      labels=c("Employment","Jobless","Inactive"))

grid.arrange(eag1, eag2, nrow = 2 , top ="Economic Activity")
```



2006년 45세~55세, 56세 ~ 65세, 66세~75세, 76세~85세, 86세~95세의 연령그룹 데이터

In [18]:

```
klosa %>% filter((year==2006 & age < 56) |
                     (year==2008 & age < 58) |
                     (year==2010 & age < 60) |
                     (year==2012 & age < 62) |
                     (year==2014 & age < 64) |
                     (year==2016 & age < 66)) %>%
    select(year,gender,age,ageg,emp,present_ecotype,earned,adl,iadl) %>% mutate(tfr="45to55") -> r1


klosa %>% filter((year==2006 & age > 55 & age < 66) |
                     (year==2008 & age > 57 & age < 68) |
                     (year==2010 & age > 59 & age < 70) |
                     (year==2012 & age > 61 & age < 72) |
                     (year==2014 & age > 63 & age < 74) |
                     (year==2016 & age > 65 & age < 76)) %>%
    select(year,gender,age,ageg,emp,present_ecotype,earned,adl,iadl) %>% mutate(tfr="56to65") -> r2


klosa %>% filter((year==2006 & age > 65 & age < 76) |
                     (year==2008 & age > 67 & age < 78) |
                     (year==2010 & age > 69 & age < 80) |
                     (year==2012 & age > 71 & age < 82) |
                     (year==2014 & age > 73 & age < 84) |
                     (year==2016 & age > 75 & age < 86)) %>%
    select(year,gender,age,ageg,emp,present_ecotype,earned,adl,iadl) %>% mutate(tfr="66to75") -> r3


klosa %>% filter((year==2006 & age > 75 & age < 86) |
                     (year==2008 & age > 77 & age < 88) |
                     (year==2010 & age > 79 & age < 90) |
                     (year==2012 & age > 81 & age < 92) |
                     (year==2014 & age > 83 & age < 94) |
                     (year==2016 & age > 85 & age < 96)) %>%
    select(year,gender,age,ageg,emp,present_ecotype,earned,adl,iadl) %>% mutate(tfr="76to85") -> r4


klosa %>% filter((year==2006 & age > 85 & age < 96) |
                     (year==2008 & age > 87 & age < 98) |
                     (year==2010 & age > 89 & age < 100) |
                     (year==2012 & age > 91 & age < 102) |
                     (year==2014 & age > 93 & age < 104) |
                     (year==2016 & age > 95 & age < 106)) %>%
    select(year,gender,age,ageg,emp,present_ecotype,earned,adl,iadl) %>% mutate(tfr="86to95") -> r5

rbind(r1,r2,r3,r4,r5) -> klosabyage
```

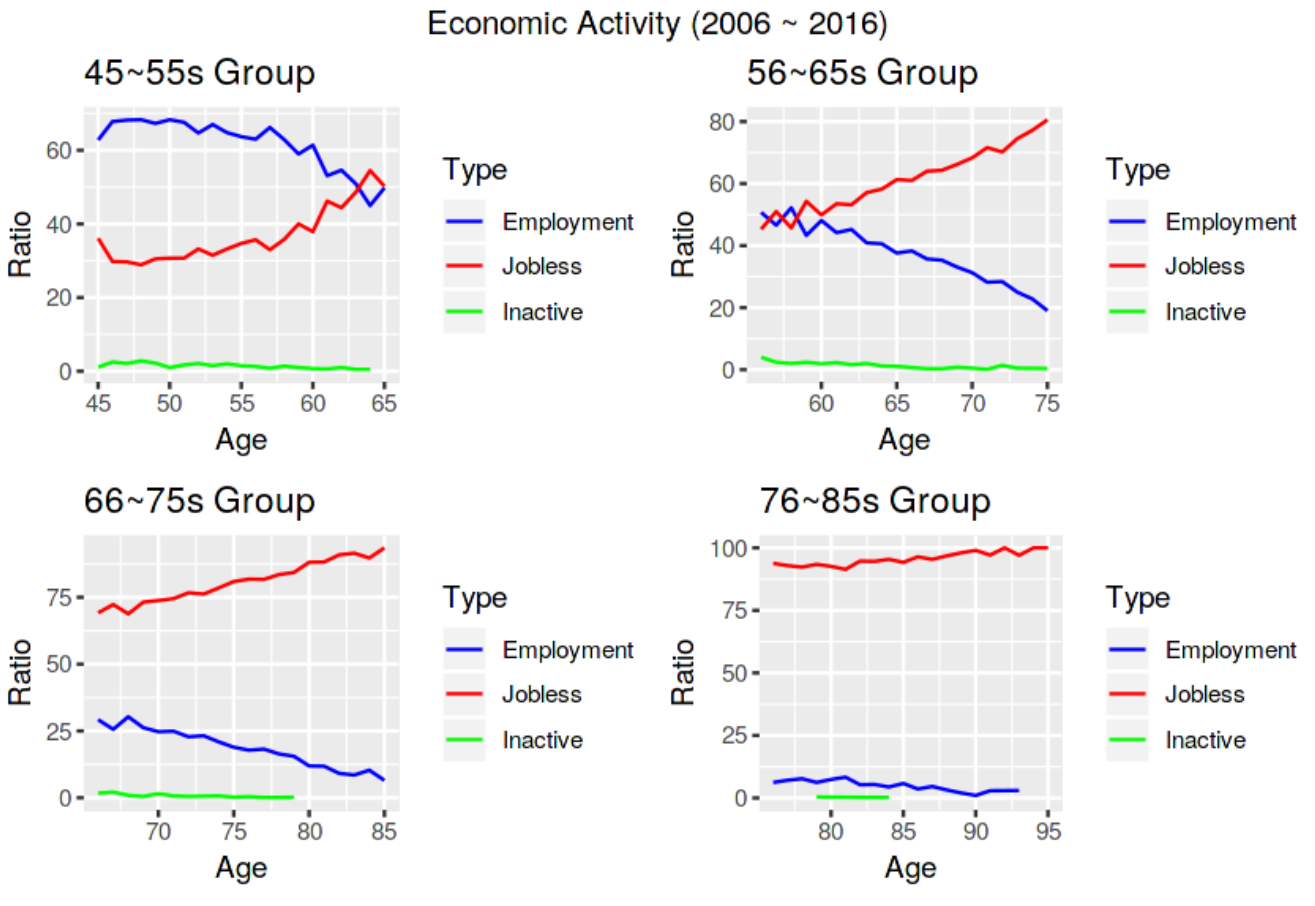
2.4 2006년 ~ 2016년 연령대별 경제활동

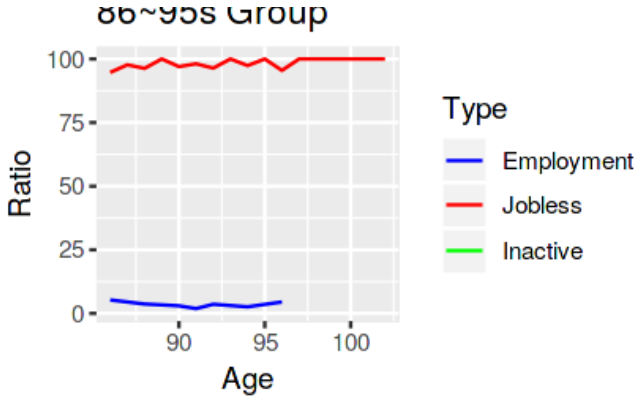
In [19]:

```
klosabyage %>% mutate(present_ecotype=ifelse(present_ecotype==1, "Employment",
                                             ifelse(present_ecotype==2,"Jobless","Inactive"))) %>%
               group_by(age,tfr,present_ecotype) %>%
               summarise(n=n()) %>%
               mutate(total=sum(n)) %>%
               mutate(pct=round(n/total*100,1))->ecoact_ageg_year
```

In [20]:

```
ggplot(ecoact_ageg_year[ecoact_ageg_year$tfr=="45to55",],aes(age,pct,color=present_ecotype))+
      geom_line() +
      labs(y="Ratio",x="Age",title="45~55s Group") +
      scale_color_manual( values=c("Blue","Red","Green"),
                          name = ("Type"),
                          labels = c("Employment","Jobless","Inactive"))-> p1
 ggplot(ecoact_ageg_year[ecoact_ageg_year$tfr=="56to65",],aes(age,pct,color=present_ecotype))+
      geom_line()+
      labs(y="Ratio",x="Age",title="56~65s Group") +
      scale_color_manual( values=c("Blue","Red","Green"),
                          name = ("Type"),
                          labels = c("Employment","Jobless","Inactive")) -> p2
 ggplot(ecoact_ageg_year[ecoact_ageg_year$tfr=="66to75",],aes(age,pct,color=present_ecotype))+
   geom_line()+
   labs(y="Ratio",x="Age",title="66~75s Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Jobless","Inactive")) -> p3
 ggplot(ecoact_ageg_year[ecoact_ageg_year$tfr=="76to85",],aes(age,pct,color=present_ecotype))+
   geom_line()+
   labs(y="Ratio",x="Age",title="76~85s Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Jobless","Inactive")) -> p4
 ggplot(ecoact_ageg_year[ecoact_ageg_year$tfr=="86to95",],aes(age,pct,color=present_ecotype))+
   geom_line()+
   labs(y="Ratio",x="Age",title="86~95s Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Jobless","Inactive")) -> p5
 grid.arrange(p1, p2, p3, p4, p5, top="Economic Activity (2006 ~ 2016)", nrow = 3)
```





2.5 2006년 ~ 2016년 연령대별 성별 경제활동

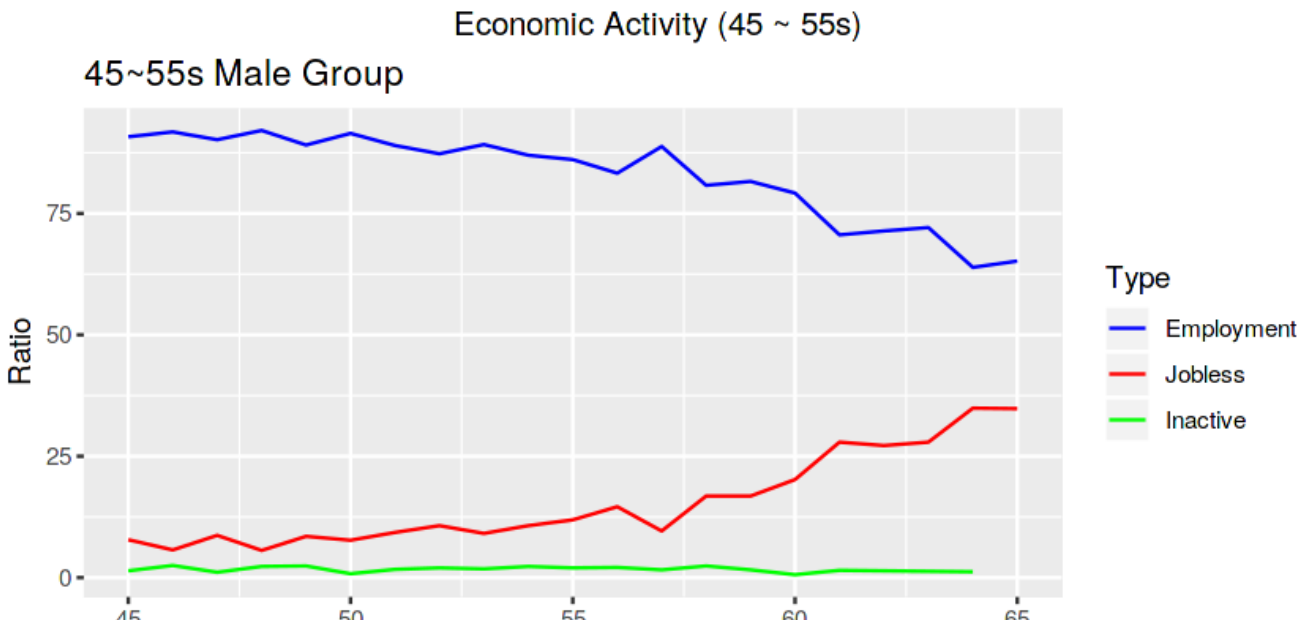
In [21]:

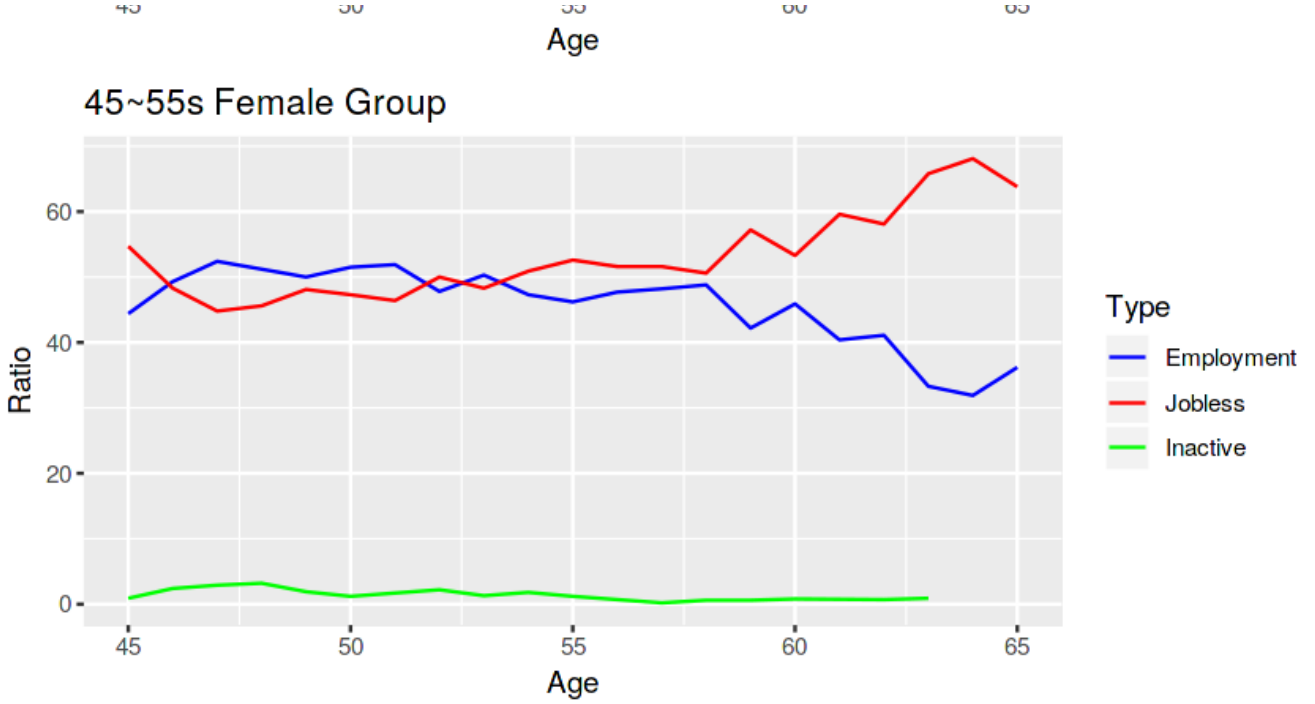
```
klosabyage %>% mutate(present_ecotype=ifelse(present_ecotype==1, "Employment",
                                             ifelse(present_ecotype==2,"Jobless","Inactive"))) %>%
   select(age,gender,tfr,present_ecotype) %>%
   group_by(age,gender,tfr,present_ecotype) %>%
   summarise(n=n()) %>%
   mutate(total=sum(n)) %>%
   mutate(pct=round(n/total*100,1)) ->ecoact_gender
```

45세~55세 , 56세~65세 그룹의 활동

In [22]:

```
ggplot(ecoact_gender[ecoact_gender$tfr=="45to55" & ecoact_gender$gender=="male",],aes(age,pct,color
=present_ecotype))+
   geom_line()+
   labs(y="Ratio",x="Age",title="45~55s Male Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Jobless","Inactive")) -> g1

 ggplot(ecoact_gender[ecoact_gender$tfr=="45to55" & ecoact_gender$gender=="female",],aes(age,pct,co
lor=present_ecotype))+
   geom_line()+
   labs(y="Ratio",x="Age",title="45~55s Female Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Jobless","Inactive")) -> g2
grid.arrange(g1, g2, top="Economic Activity (45 ~ 55s)", nrow = 2)
```



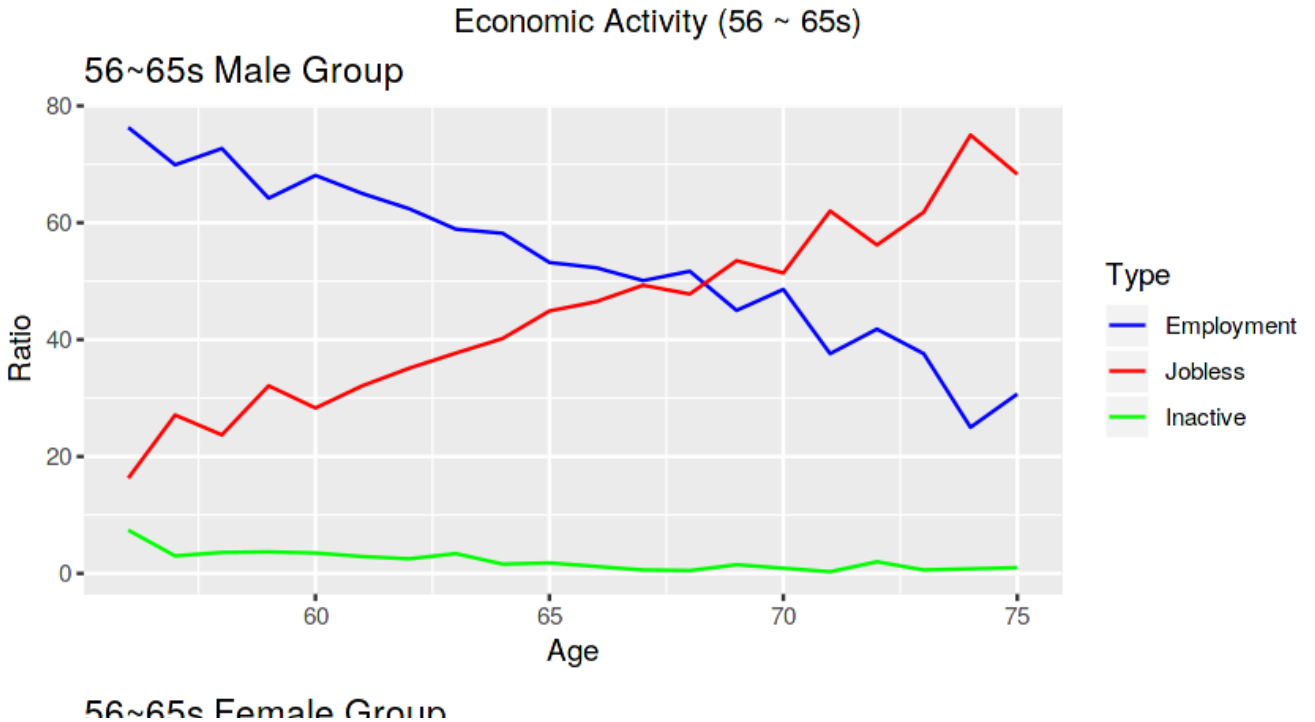


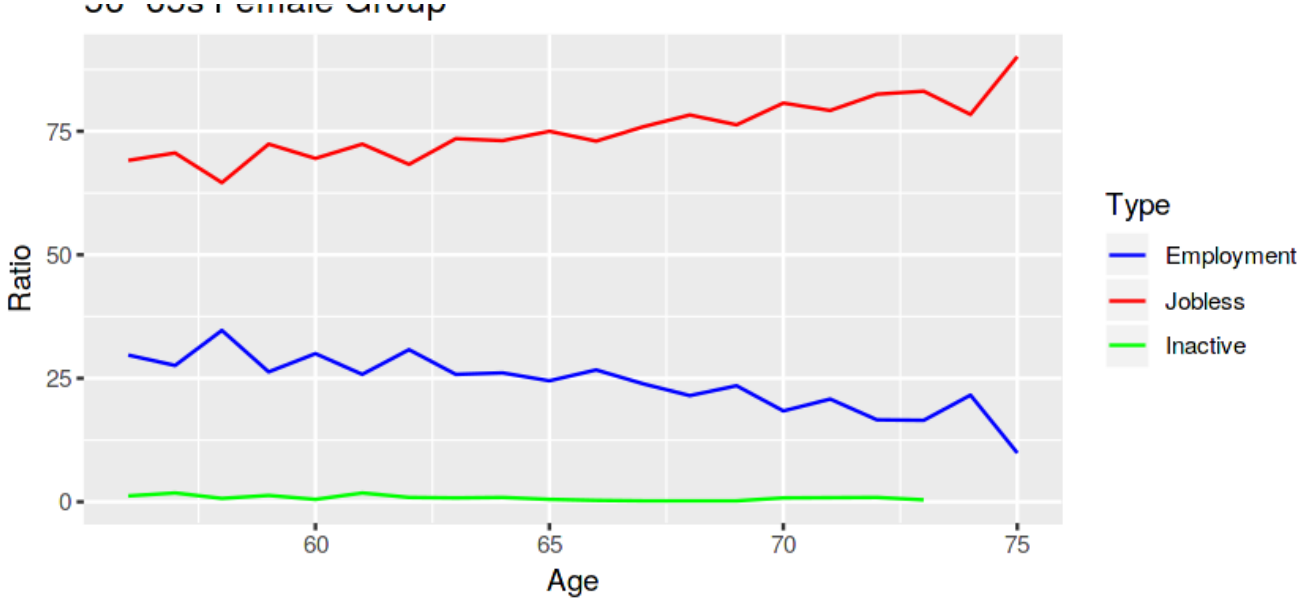
In [23]:

```
ggplot(ecoact_gender[ecoact_gender$tfr=="56to65" & ecoact_gender$gender=="male",],aes(age,pct,color
=present_ecotype))+
   geom_line()+
   labs(y="Ratio",x="Age",title="56~65s Male Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Jobless","Inactive")) -> g3
```

In [24]:

```
ggplot(ecoact_gender[ecoact_gender$tfr=="56to65" & ecoact_gender$gender=="female",],aes(age,pct,col
or=present_ecotype))+
   geom_line()+
   labs(y="Ratio",x="Age",title="56~65s Female Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Jobless","Inactive")) -> g4
grid.arrange(g3, g4, top="Economic Activity (56 ~ 65s)", nrow = 2)
```





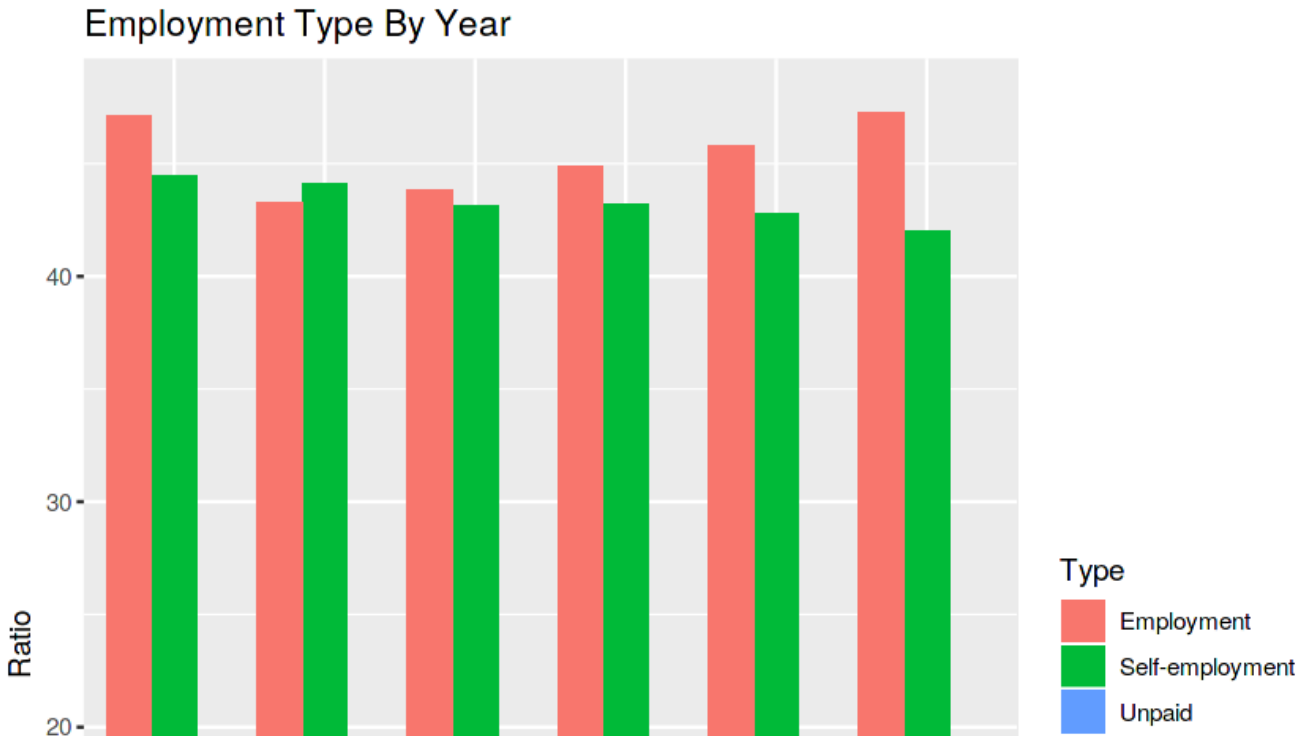
2.6 연도별 일자리 고용형태 [임금고용, 개인사업, 무급(가족,친척의 일을 돕는다)]

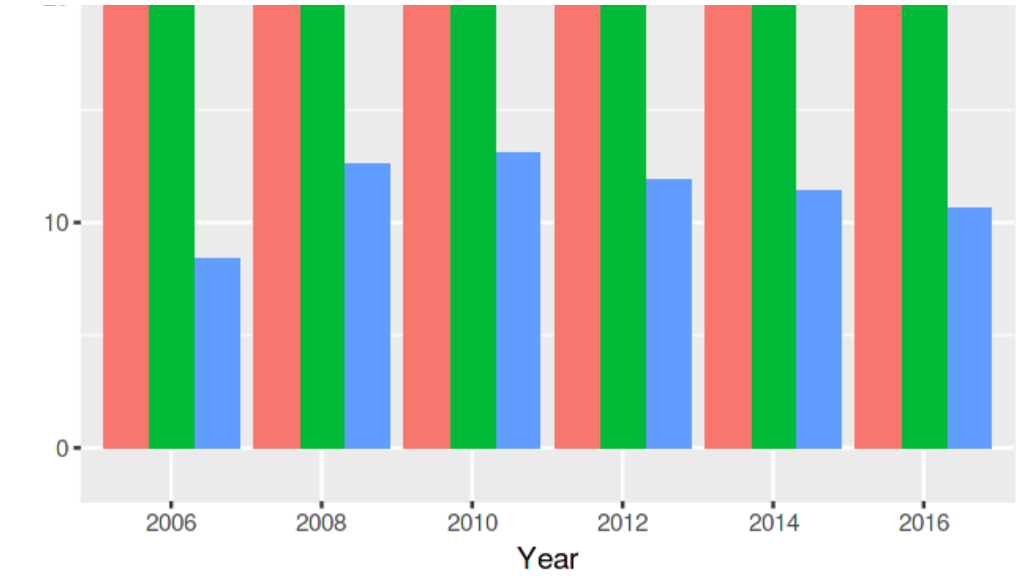
In [25]:

```
klosa %>% filter(!is.na(emp)&!is.na(year)) %>%
  select(year,emp) %>%
  mutate(emps=ifelse(emp==3|emp==4,3,emp)) %>%
  group_by(year,emps) %>%
  summarise(n=n()) %>%
  mutate(total=sum(n)) %>%
  mutate(pct=round(n/total*100,1))->ecoact_emp_year
```

In [26]:

```
ggplot(ecoact_emp_year,aes(as.factor(year), pct ,fill=as.factor(emps))) +
  geom_col(position='dodge')+
  labs(y="Ratio",x="Year",title="Employment Type By Year") +
  theme_grey() +
  scale_fill_discrete(name="Type",
                      labels=c("Employment","Self-employment","Unpaid"))
```





2.7 2006년 ~ 2016년 연령별 일자리 고용형태

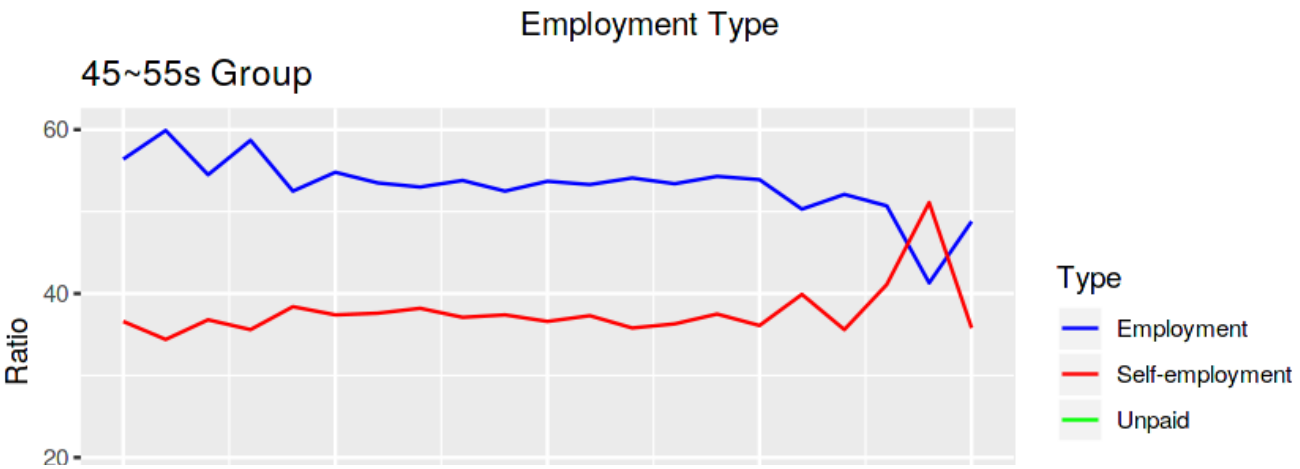
In [27]:

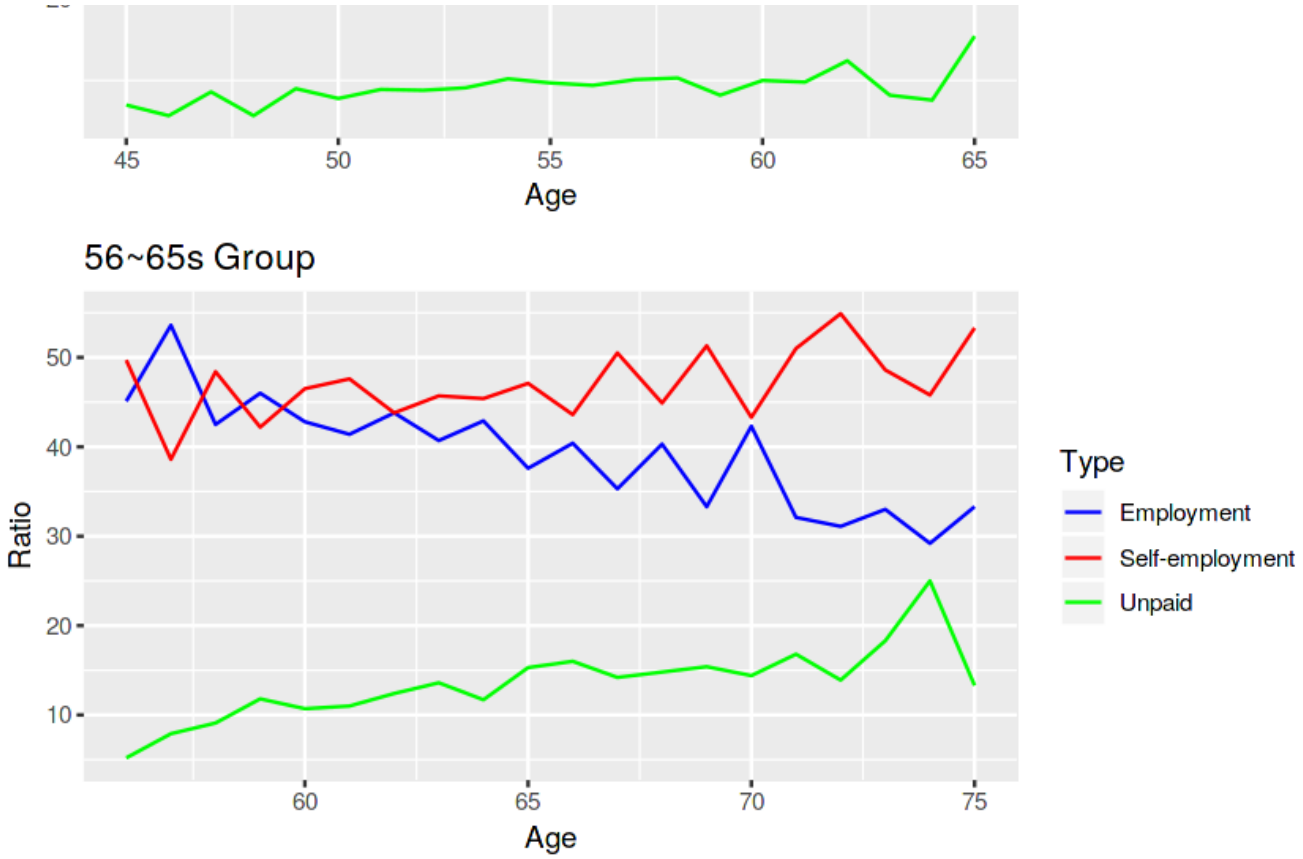
```
klosabyage %>% filter(!is.na(emp)&!is.na(year)) %>%
   select(age,tfr,emp) %>%
   mutate(emps=ifelse(emp==3|emp==4,3,emp)) %>%
   group_by(age,tfr,emps) %>%
   summarise(n=n()) %>%
   mutate(total=sum(n)) %>%
   mutate(pct=round(n/total*100,1))->ecoact_emp_tfr
```

45세~55세 , 56세~65세 그룹의 고용형태

In [28]:

```
ggplot(ecoact_emp_tfr[ecoact_emp_tfr$tfr=="45to55",],aes(age,pct,color=factor(emps)))+
   geom_line()+
   labs(y="Ratio",x="Age",title="45~55s Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Self-employment","Unpaid")) -> e1

 ggplot(ecoact_emp_tfr[ecoact_emp_tfr$tfr=="56to65",],aes(age,pct,color=factor(emps)))+
   geom_line()+
   labs(y="Ratio",x="Age",title="56~65s Group") +
   scale_color_manual( values=c("Blue","Red","Green"),
                       name = ("Type"),
                       labels = c("Employment","Self-employment","Unpaid")) -> e2

 grid.arrange(e1, e2, top="Employment Type ", nrow = 2)
```





3. 소득

3.1 연도별 근로소득

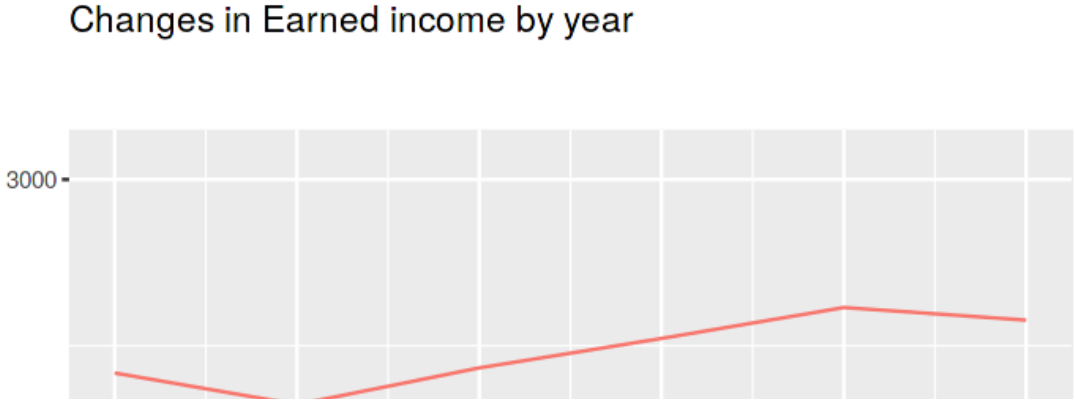
(2006년 45세~55세, 56세 ~ 65세, 66세~75세, 76세~85세, 86세~95세의 연령그룹)

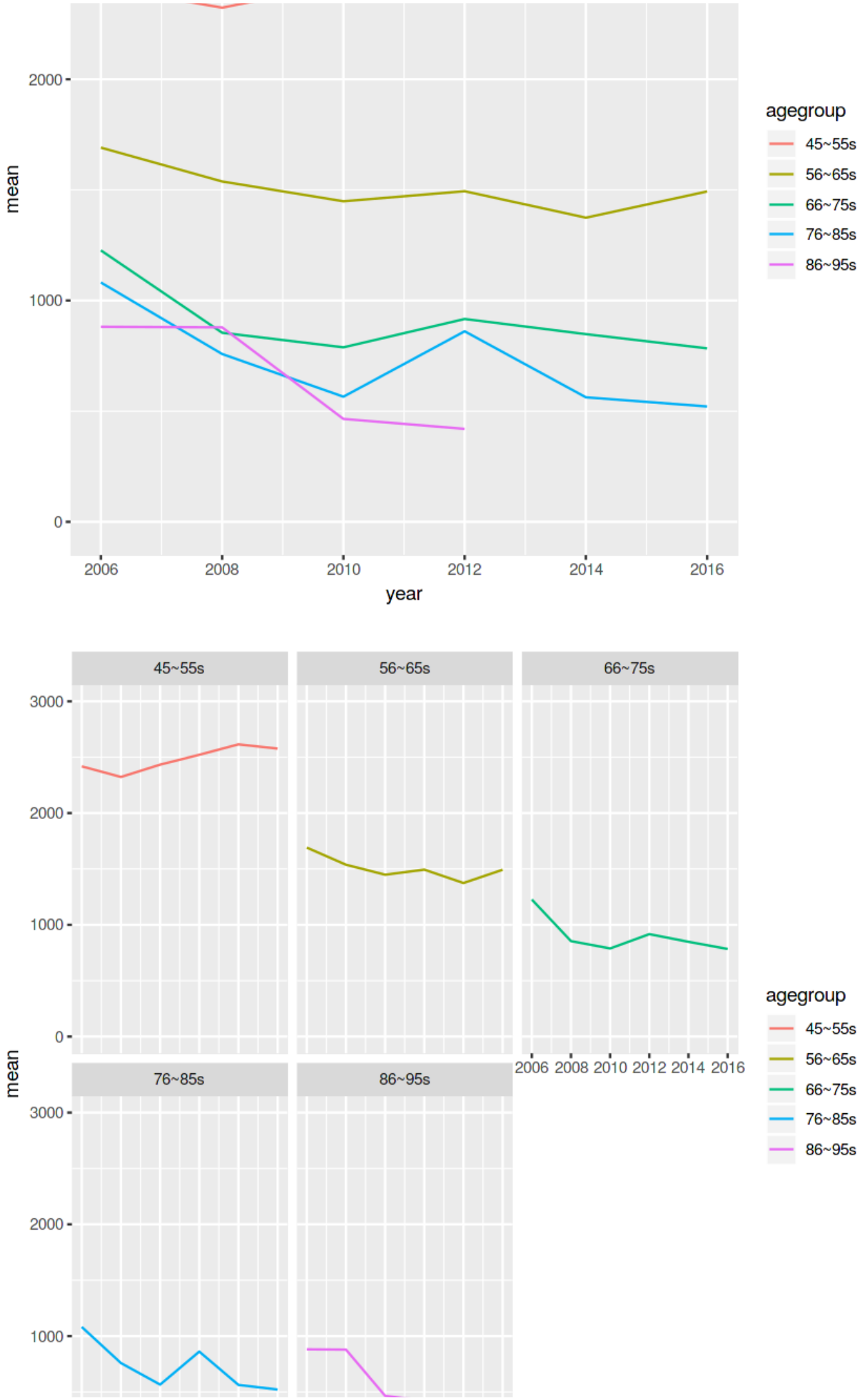
In [29]:

```
klosabyage %>% select(year,tfr,earned)%>%
               filter(!is.na(earned)) %>%
               mutate(tfr=ifelse(tfr=="45to55","45~55s",
                          ifelse(tfr=="56to65","56~65s",
                          ifelse(tfr=="66to75","66~75s",
                          ifelse(tfr=="76to85","76~85s", "86~95s"))))) %>%
               rename(agegroup=tfr) %>%
               group_by(year,agegroup) %>%
               summarise(mean = mean(earned, na.rm=T)) -> ep1
```

In [30]:

```
ep1 %>% ggplot(aes(year,mean,color=agegroup))+geom_line()+ylim(0,3000)+ggtitle("Changes in Earned
income by year\n\n")
ep1 %>% ggplot(aes(year,mean,color=agegroup))+geom_line()+facet_wrap(~ agegroup)+ylim(0,3000)
```

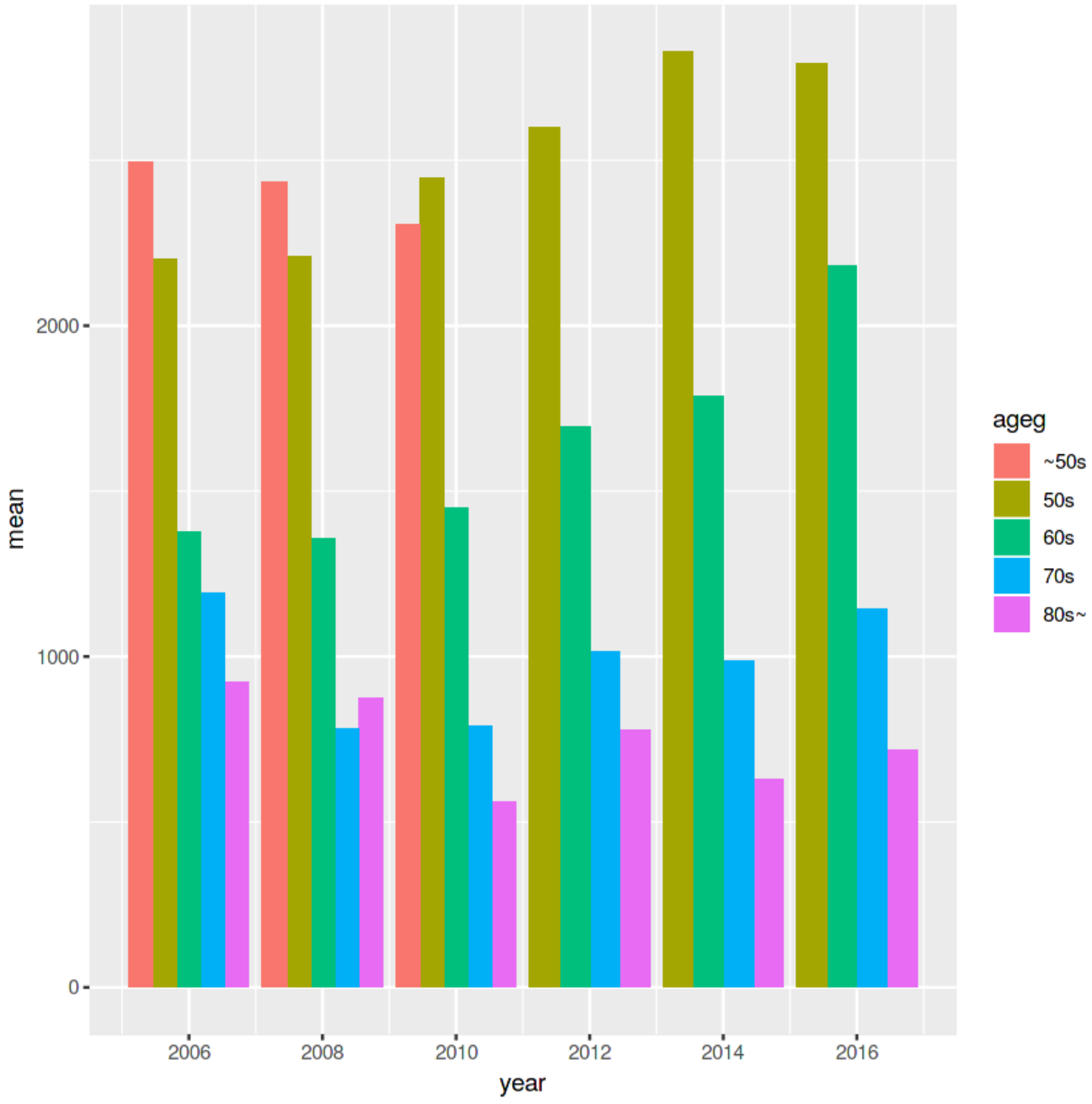



mean
2000
1000
0
2006
2008
2010
2012
2014
2016
year
agegroup
45~55s
56~65s
66~75s
76~85s
86~95s
45~55s
56~65s
66~75s
3000
2000
1000
0
2006 2008 2010 2012 2014 2016
76~85s
86~95s
3000
2000
1000
mean
agegroup
45~55s
56~65s
66~75s
76~85s
86~95s

3.2 연도별 연령대의 소득

In [31]:

```
klosa %>%
    filter(!is.na(earned)) %>%
    select(year,ageg,earned) %>%
    group_by(year,ageg) %>%
    summarise(mean=mean(earned)) %>%
    ggplot(aes(year,mean,fill=ageg))+geom_col(position="dodge") +
    scale_x_continuous(breaks=c(2006,2008,2010,2012,2014,2016))
```



3.3 연도에 따른 연령별 근로소득의 변화

In [32]:

```
klosa %>%
filter(!is.na(earned)) %>%
filter(ageg=="~50s") %>%
select(year,ageg,earned) %>%
group_by(year) %>%
summarise("~50"=mean(earned)) ->f

klosa %>%
filter(!is.na(earned)) %>%
filter(ageg=="50s") %>%
select(year,ageg,earned) %>%
group_by(year) %>%
summarise("50s"=mean(earned)) ->g


klosa %>%
filter(!is.na(earned)) %>%
filter(ageg=="60s") %>%
select(year,ageg,earned) %>%
group_by(year) %>%
summarise("60s"=mean(earned)) ->h

klosa %>%
filter(!is.na(earned)) %>%
filter(ageg=="70s") %>%
select(year,ageg,earned) %>%
group_by(year) %>%
summarise("70s"=mean(earned)) ->i

klosa %>%
filter(!is.na(earned)) %>%
filter(ageg=="80s~") %>%
select(year,ageg,earned) %>%
group_by(year) %>%
summarise("80s~"=mean(earned)) ->j
```

In [33]:

```
right_join(f,g) %>% left_join(h) %>% left_join(i) %>% left_join(j) ->fghij
fghij
```

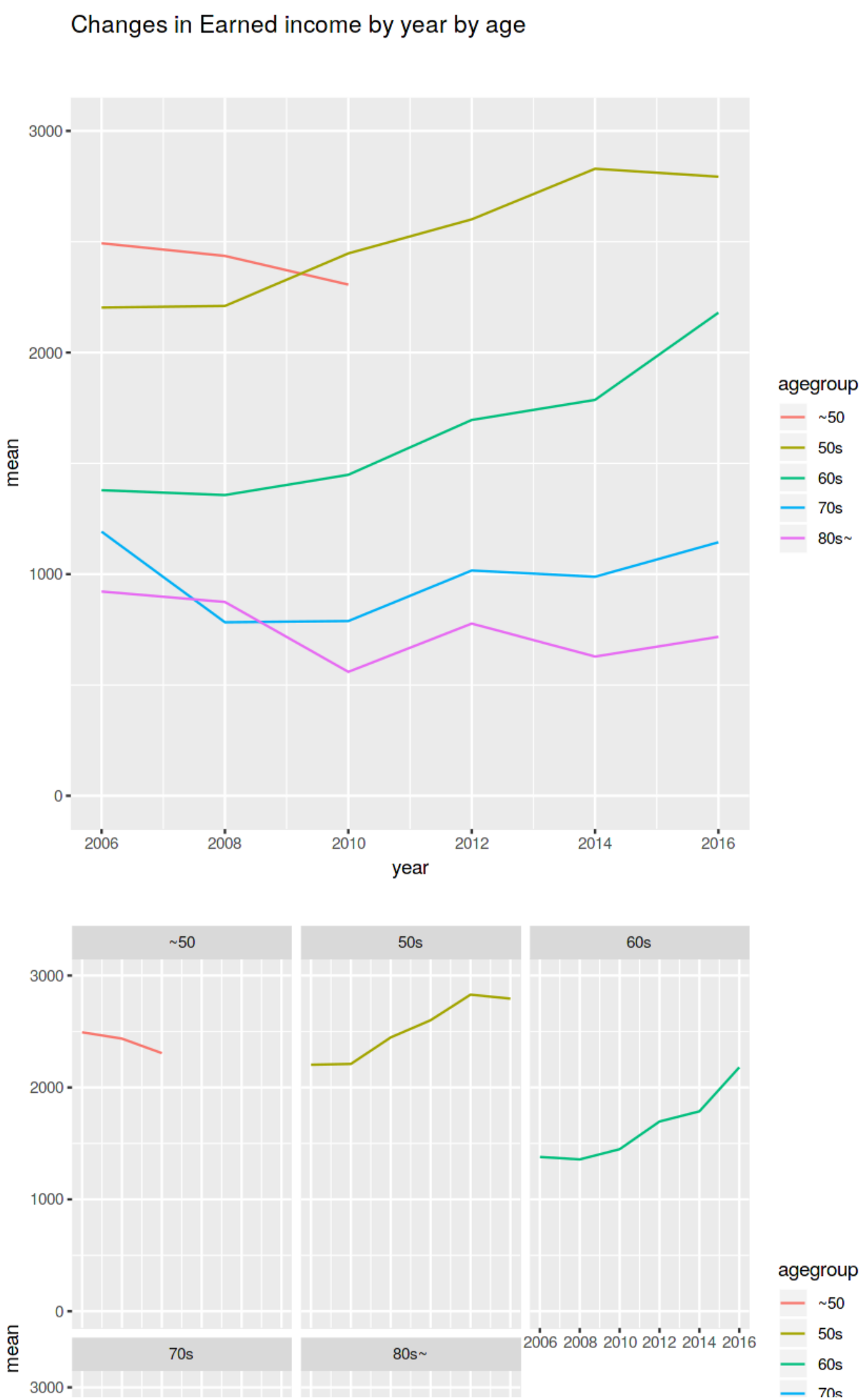
```
Joining, by = "year"
Joining, by = "year"
Joining, by = "year"
Joining, by = "year"
```

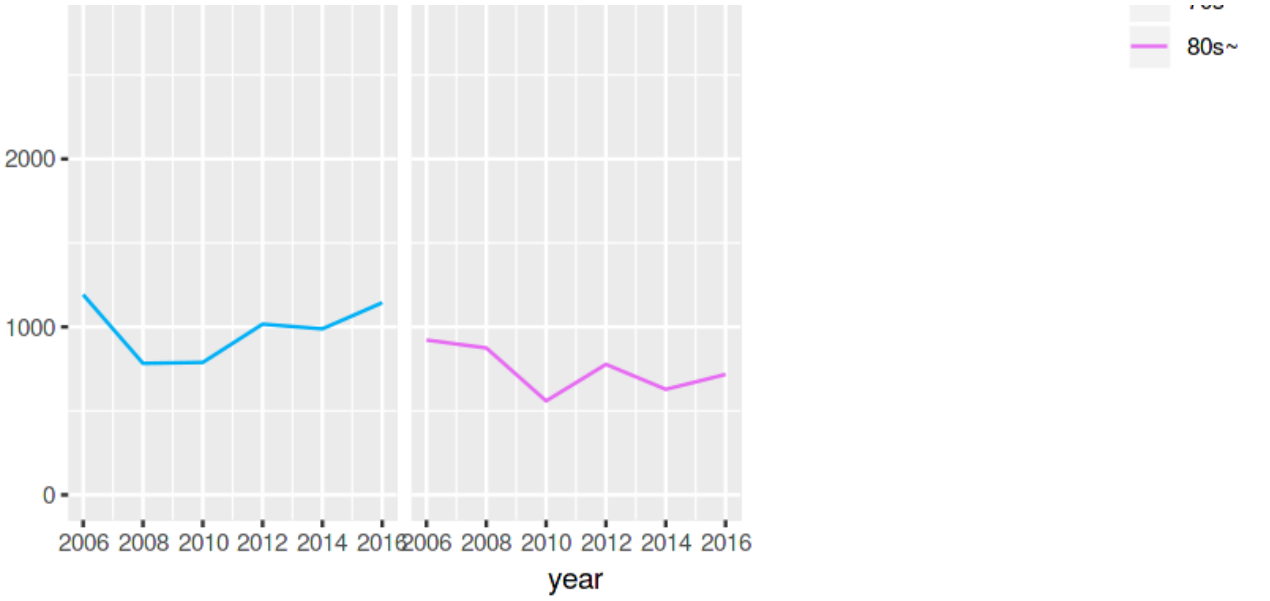
year	~50	50s	60s	70s	80s~
2006	2492.911	2203.051	1378.692	1191.8818	921.6316
2008	2435.997	2210.155	1356.993	782.8668	874.4355
2010	2306.345	2447.349	1448.460	788.4990	559.1562
2012	NA	2600.991	1695.823	1016.3766	777.0139
2014	NA	2829.167	1786.510	988.3170	628.2128
2016	NA	2793.422	2179.928	1144.0840	717.2806

In [34]:

```
gather(fghij, key="agegroup", value="mean",`~50`,`50s`,`60s`,`70s`,`80s~`) -> gathering2

gathering2 %>% ggplot(aes(year,mean,color=agegroup))+geom_line()+ylim(0,3000)+ggtitle("Changes in
Earned income by year by age\n\n")
gathering2 %>% ggplot(aes(year,mean,color=agegroup))+geom_line()+facet_wrap(~ agegroup)+ylim(0,3000
)
```

```
Warning message:
“Removed 3 rows containing missing values (geom_path).”Warning message:
“Removed 3 rows containing missing values (geom_path).”
```

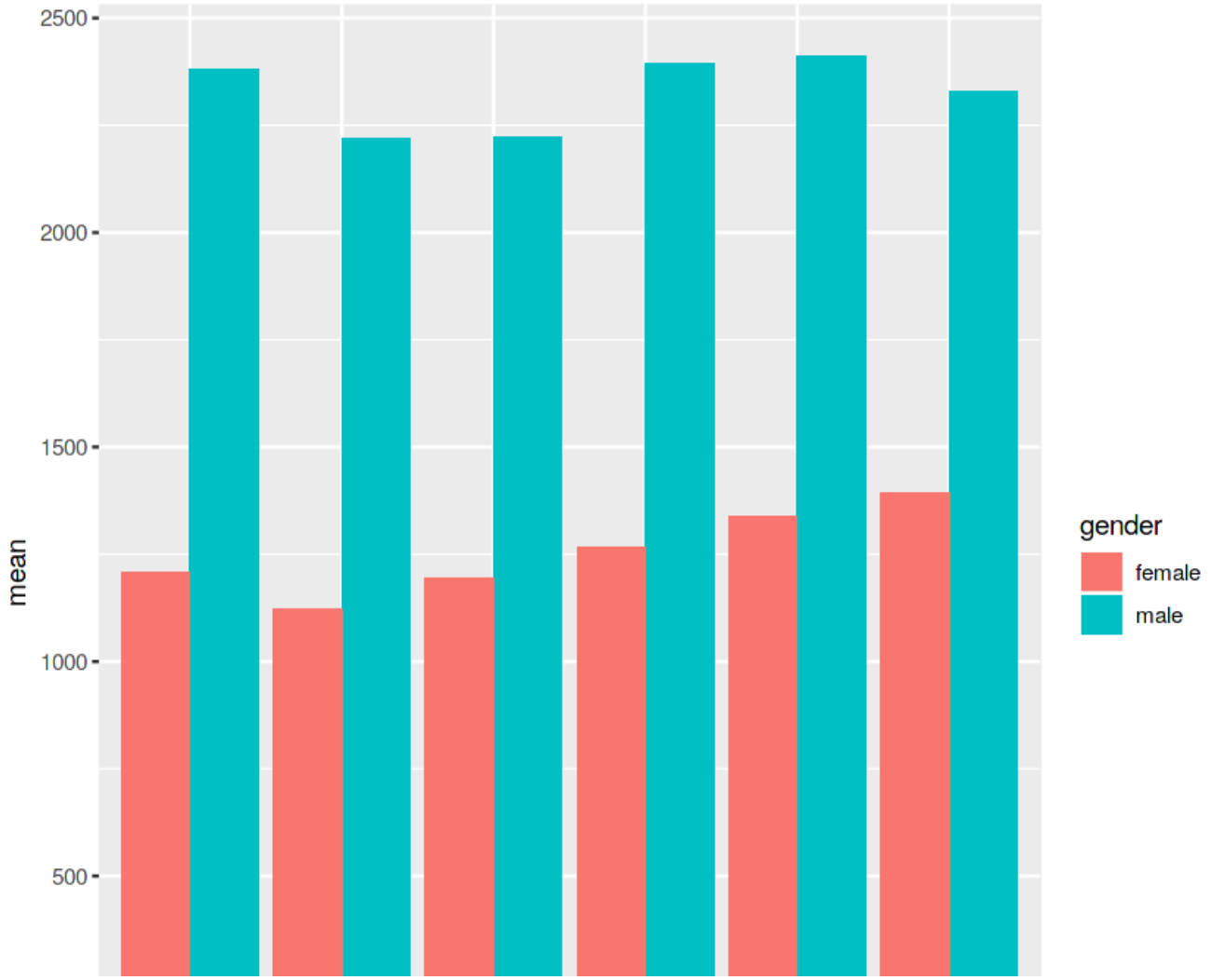

Changes in Earned income by year by age
mean
year
3000
2000
1000
0
2006
2008
2010
2012
2014
2016
agegroup
~50
50s
60s
70s
80s~
~50
50s
60s
70s
80s~

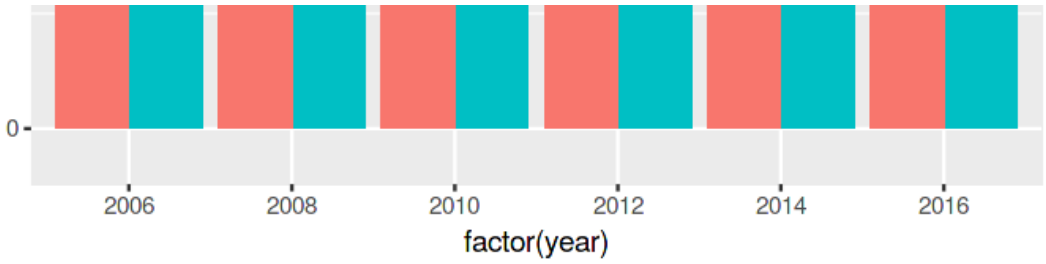


3.4 연도별 성별 근로소득

In [35]:

```
klosa %>%
    filter(!is.na(earned)) %>%
    select(year,gender,earned) %>%
    group_by(year,gender) %>%
    summarise(mean=mean(earned)) -> meanbygender

meanbygender %>% ggplot(aes(factor(year),mean,fill=gender))+geom_col(position="dodge")
```





3.5 연도별 성별 소득의 차이(남성의 소득 - 여성의 소득)

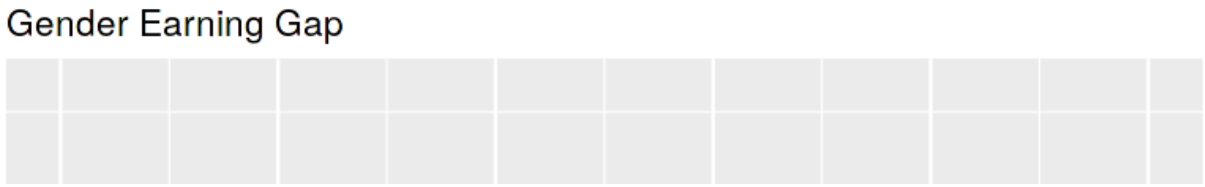
In [36]:

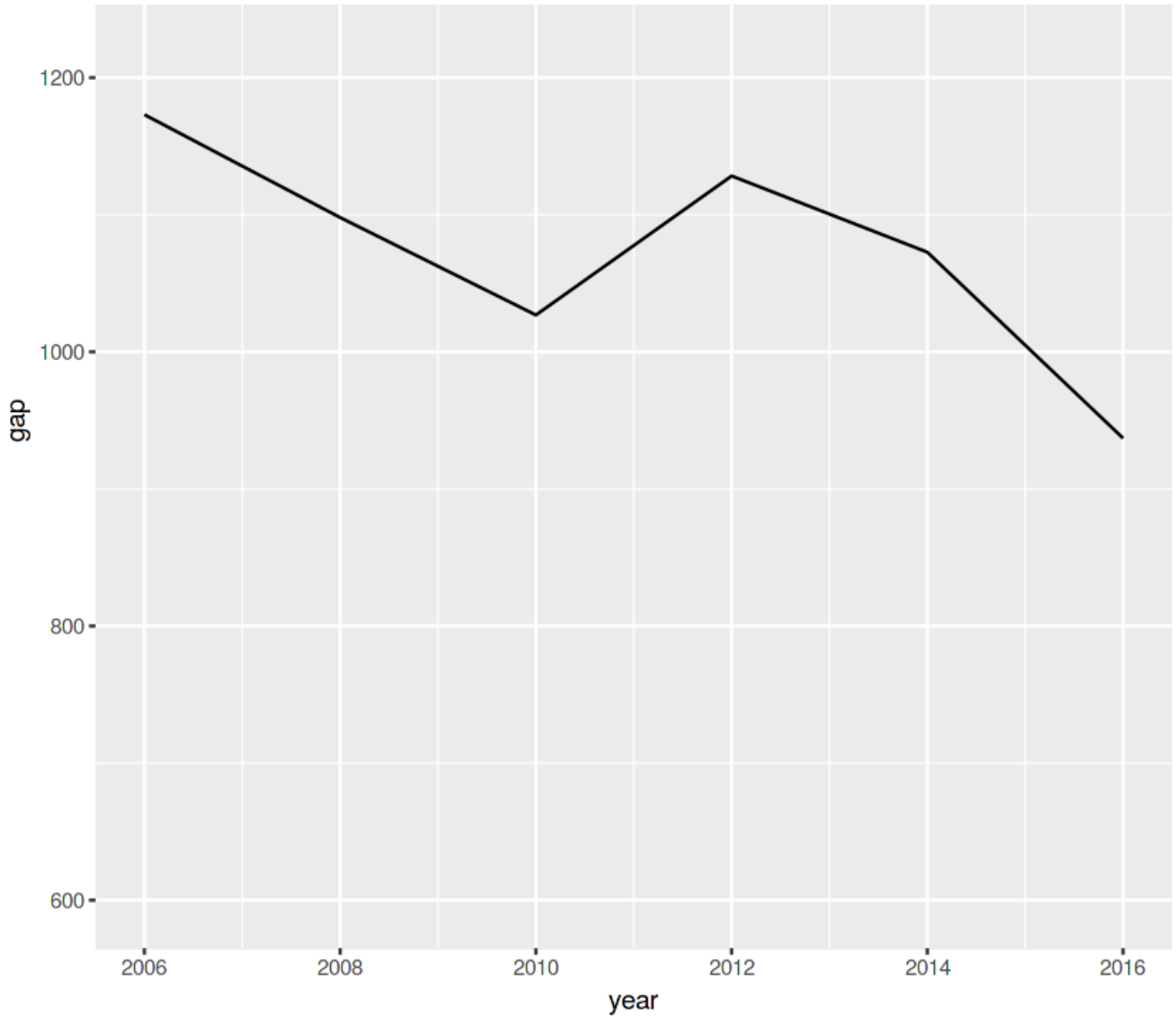
```
library(sqldf)
sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from meanbygender where year==2006 and gender=='female') x,
            (select year, mean from meanbygender where year==2006 and gender=='male') y
      where x.year=y.year") -> gap2006

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from meanbygender where year==2008 and gender=='female') x,
            (select year, mean from meanbygender where year==2008 and gender=='male') y
      where x.year=y.year") -> gap2008

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from meanbygender where year==2010 and gender=='female') x,
            (select year, mean from meanbygender where year==2010 and gender=='male') y
      where x.year=y.year") -> gap2010

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from meanbygender where year==2012 and gender=='female') x,
            (select year, mean from meanbygender where year==2012 and gender=='male') y
      where x.year=y.year") -> gap2012

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from meanbygender where year==2014 and gender=='female') x,
            (select year, mean from meanbygender where year==2014 and gender=='male') y
      where x.year=y.year") -> gap2014

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from meanbygender where year==2016 and gender=='female') x,
            (select year, mean from meanbygender where year==2016 and gender=='male') y
      where x.year=y.year") -> gap2016
```

In [37]:

```
(rbind(gap2006,gap2008,gap2010,gap2012,gap2014,gap2016) -> gendergap)
gendergap %>%
    ggplot(aes(year,gap))+geom_line()+
    scale_x_continuous(breaks=c(2006,2008,2010,2012,2014,2016))+
    ggtitle("Gender Earning Gap") +ylim(600,1300)
```

year	gap
2006	1173.0119
2008	1097.9697
2010	1026.8632
2012	1128.2876
2014	1072.6811
2016	937.0045



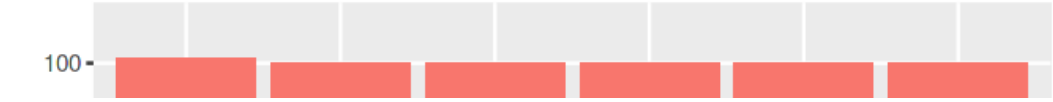


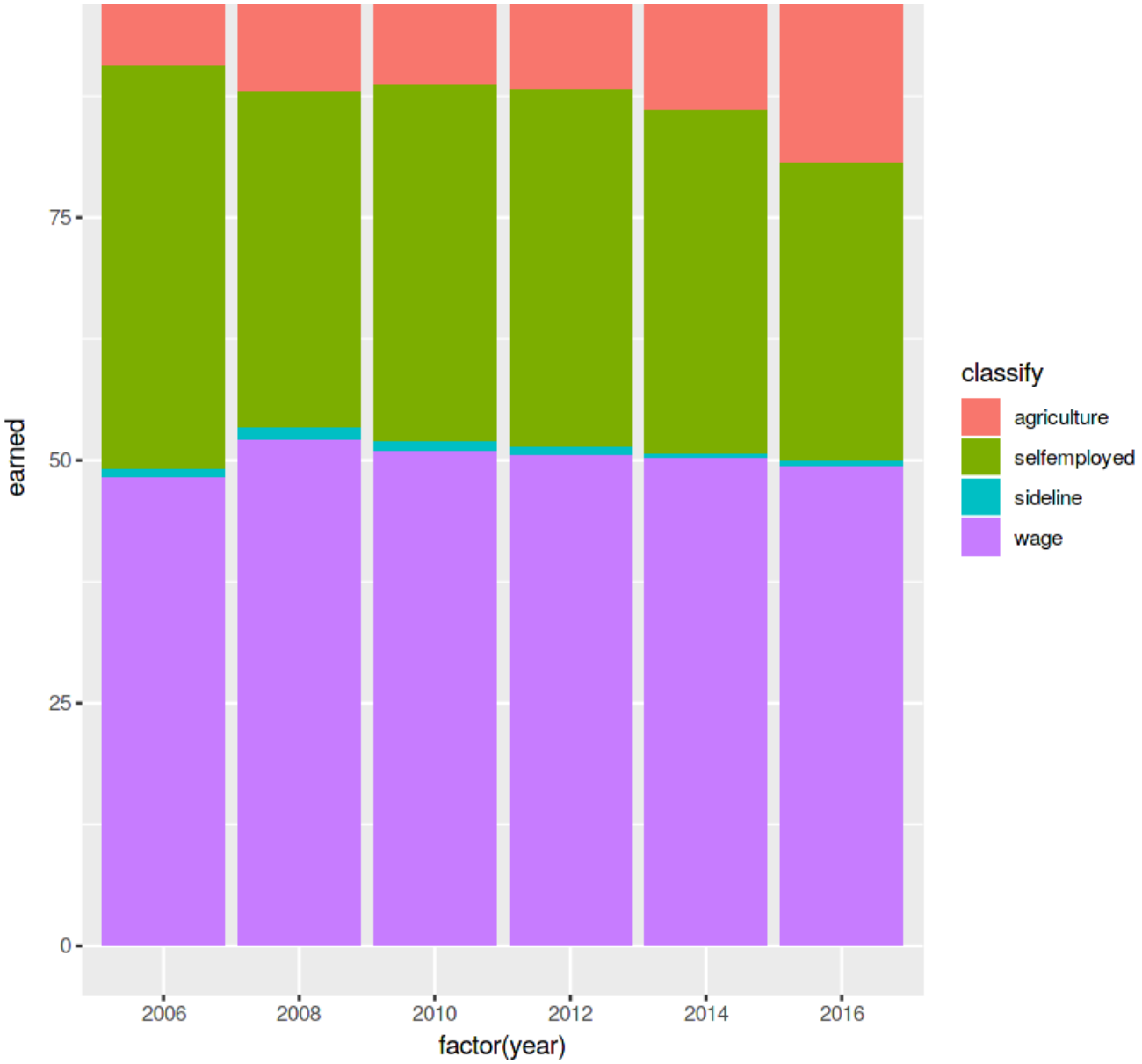
3.6 연도별 근로소득의 구성 (임금,자영업, 농어업, 부업)

- 근로총소득=임금소득+자영업소득+농어업소득+부업소득
  - wage : 임금소득
  - selfemployed : 자영업(사업체운영) 소득
  - agriculture : 농어업 소득
  - sideline : 부업소득

In [38]:

```
klosa %>%
    select(year,earned,wage,selfemployed,agriculture,sideline) %>%
    group_by(year) %>%
    summarize(earned=sum(earned,na.rm=T),
              wage=sum(wage,na.rm=T),
              selfemployed=sum(selfemployed,na.rm=T),
              agriculture=sum(agriculture,na.rm=T),
              sideline=sum(sideline,na.rm=T)) %>%
    mutate(wage=wage/earned*100) %>%
    mutate(selfemployed=selfemployed/earned*100) %>%
    mutate(agriculture=agriculture/earned*100) %>%
    mutate(sideline=sideline/earned*100) -> www

www_long=gather(www, key="classify", value="earned",`wage`,`selfemployed`,`agriculture`,`sideline`)
www_long %>% ggplot(aes(factor(year),earned,fill=classify))+geom_col()
```



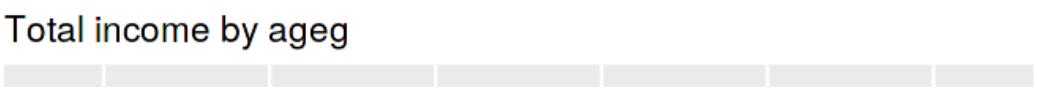


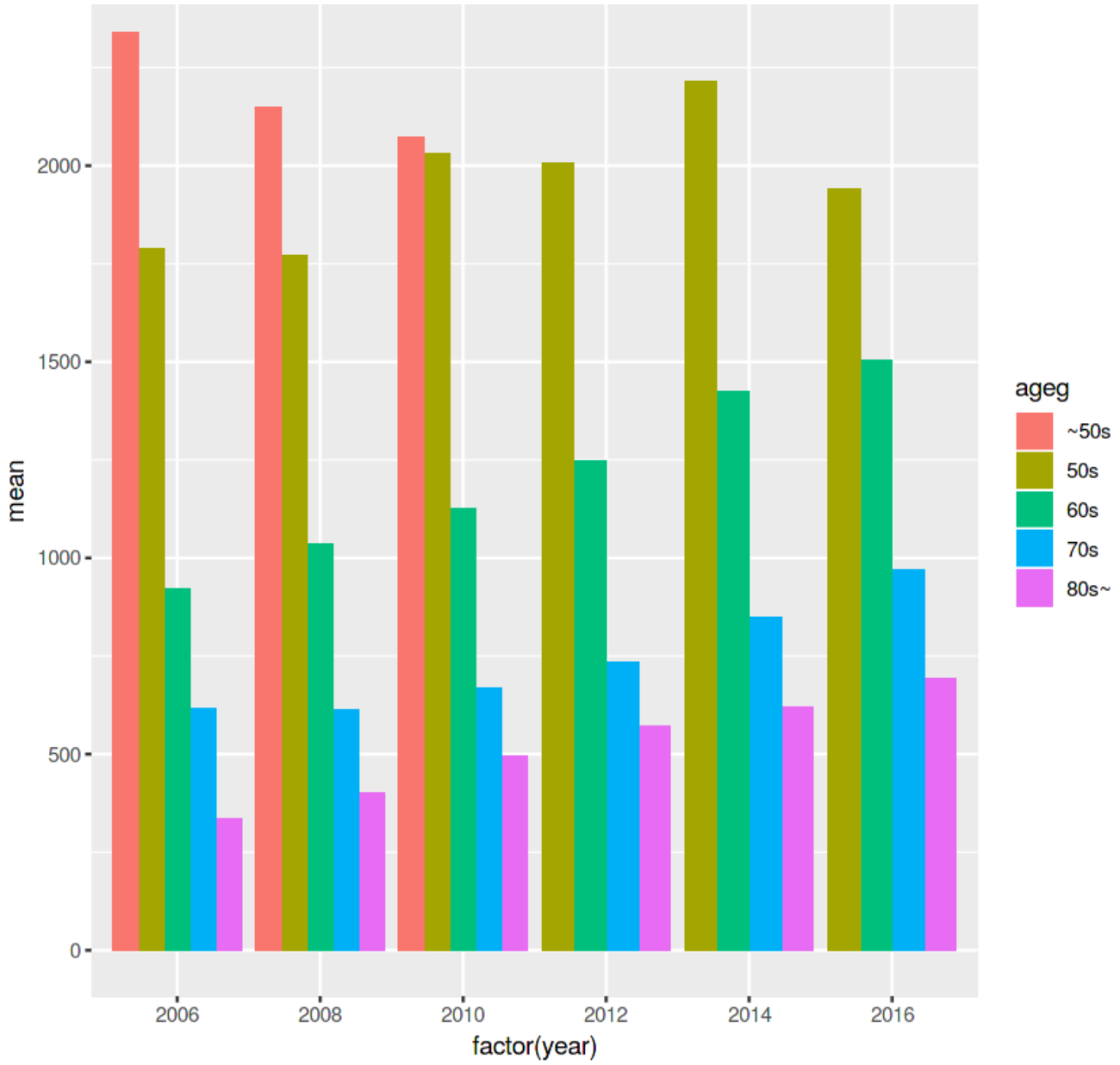
3.7 연도별 연령별 총소득 / 연도별 성별 총소득

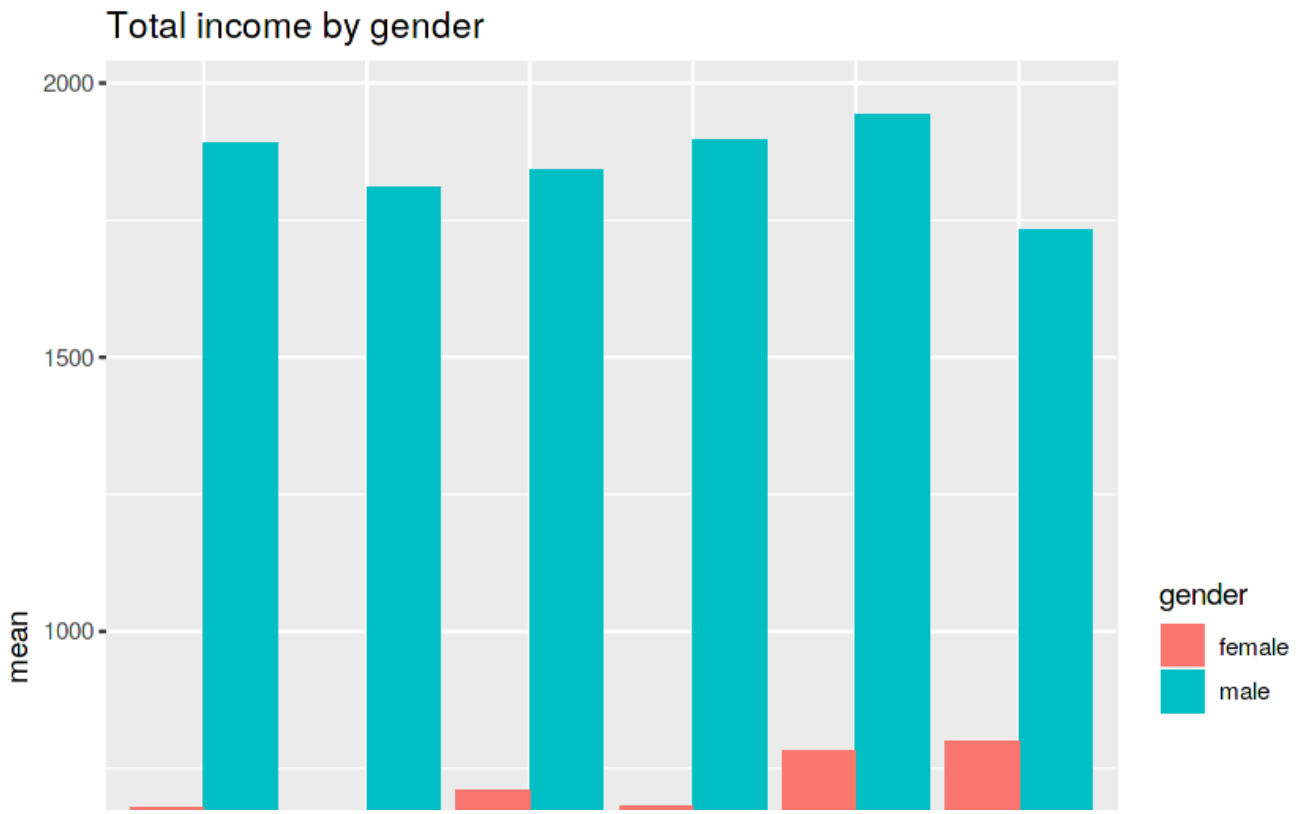
- 개인총소득(근로총소득+자산총소득+공적이전총소득+사적이전총소득+개인연금소득+ 기타소득)

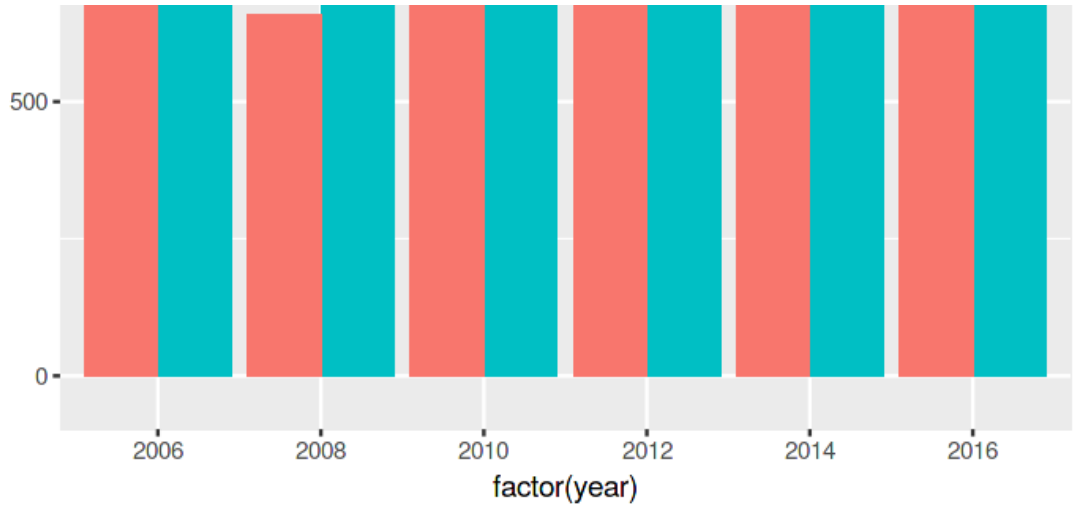
In [39]:

```
(klosa %>%
    filter(!is.na(p_inc)) %>%
    select(year,ageg,p_inc) %>%
    group_by(year,ageg) %>%
    summarise(mean = mean(p_inc)) %>%
    ggplot(aes(factor(year),mean,fill=ageg))+geom_col(position="dodge") +
    ggtitle("Total income by ageg")-> p_incbyyear)

klosa %>%
    filter(!is.na(p_inc)) %>%
    select(year,gender,p_inc) %>%
    group_by(year,gender) %>%
    summarise(mean=mean(p_inc)) -> p_incbygender

p_incbygender %>%
    ggplot(aes(factor(year),mean,fill=gender))+geom_col(position="dodge") +
    ggtitle("Total income by gender")
```




2000
1500
1000
500
0
mean
2006
2008
2010
2012
2014
2016
factor(year)
ageg
~50s
50s
60s
70s
80s~


Total income by gender
2000
1500
1000
mean
gender
female
male

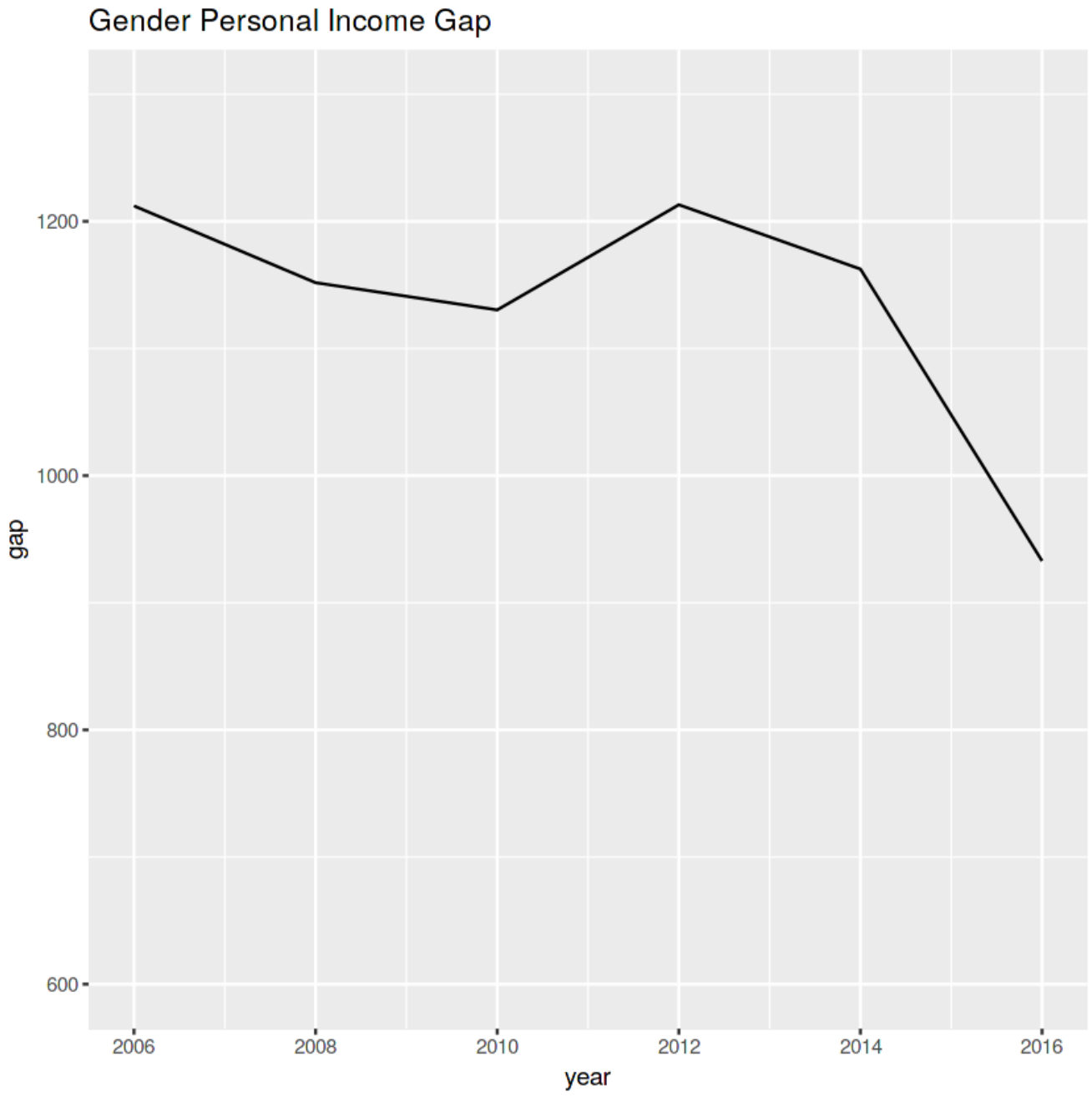


3.8 성별 총소득의 차이(남자의 총소득 - 여자의 총소득)

In [40]:

```
library(sqldf)
sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from p_incbygender where year==2006 and gender=='female') x,
            (select year, mean from p_incbygender where year==2006 and gender=='male') y
      where x.year=y.year") -> pgap2006

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from p_incbygender where year==2008 and gender=='female') x,
            (select year, mean from p_incbygender where year==2008 and gender=='male') y
      where x.year=y.year") -> pgap2008

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from p_incbygender where year==2010 and gender=='female') x,
            (select year, mean from p_incbygender where year==2010 and gender=='male') y
      where x.year=y.year") -> pgap2010

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from p_incbygender where year==2012 and gender=='female') x,
            (select year, mean from p_incbygender where year==2012 and gender=='male') y
      where x.year=y.year") -> pgap2012

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from p_incbygender where year==2014 and gender=='female') x,
            (select year, mean from p_incbygender where year==2014 and gender=='male') y
      where x.year=y.year") -> pgap2014

sqldf("select x.year, abs(x.mean-y.mean) as gap from
            (select year, mean from p_incbygender where year==2016 and gender=='female') x,
            (select year, mean from p_incbygender where year==2016 and gender=='male') y
      where x.year=y.year") -> pgap2016


(rbind(pgap2006,pgap2008,pgap2010,pgap2012,pgap2014,pgap2016) -> gendergap2)
gendergap2 %>%
    ggplot(aes(year,gap))+geom_line()+
    scale_x_continuous(breaks=c(2006,2008,2010,2012,2014,2016))+
    ggtitle("Gender Personal Income Gap") +ylim(600,1300)
```

year	gap
2006	1212.1318
2008	1151.7313
2010	1130.3365
2012	1213.0437
2014	1162.4548
2016	932.8697

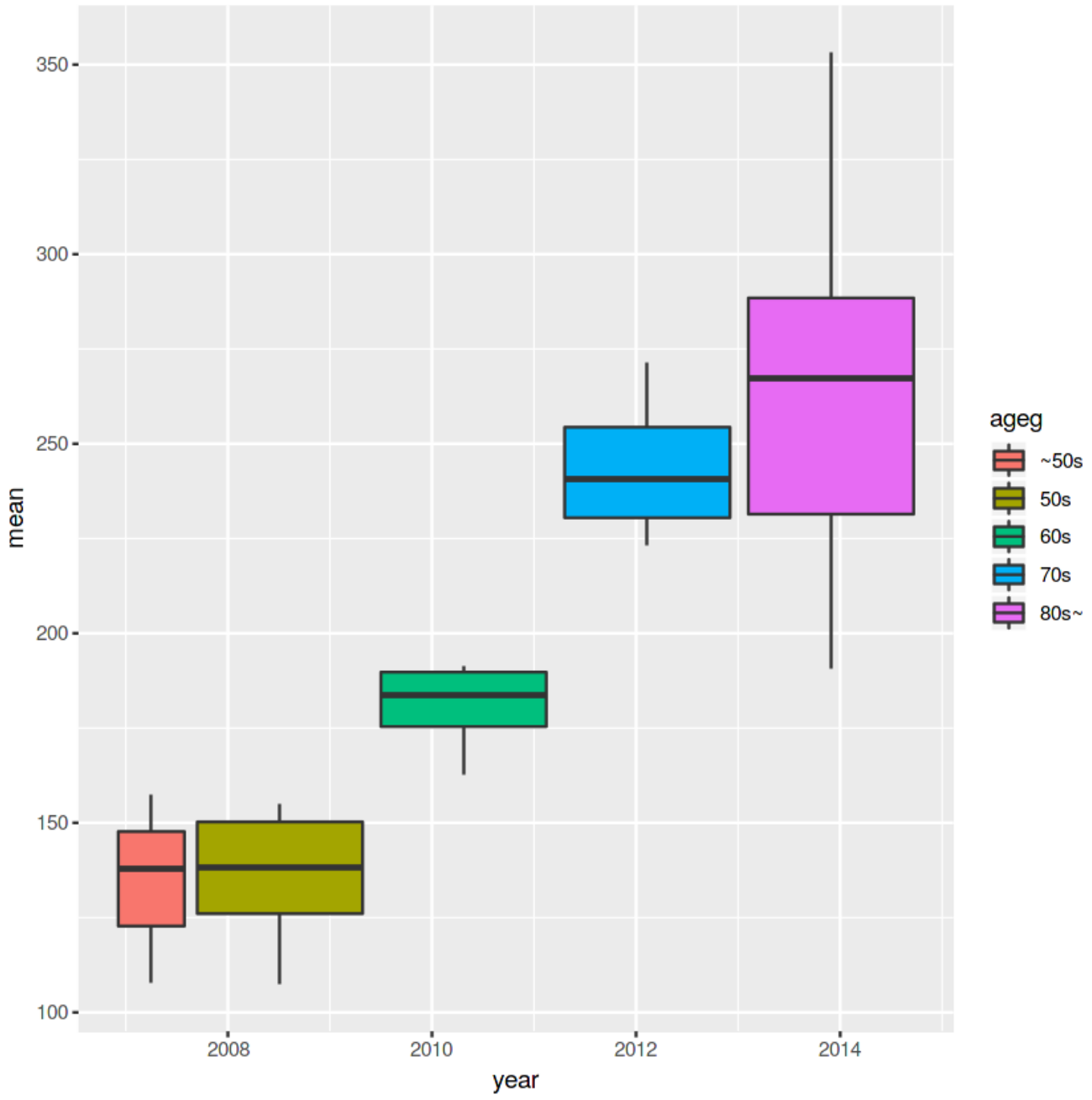


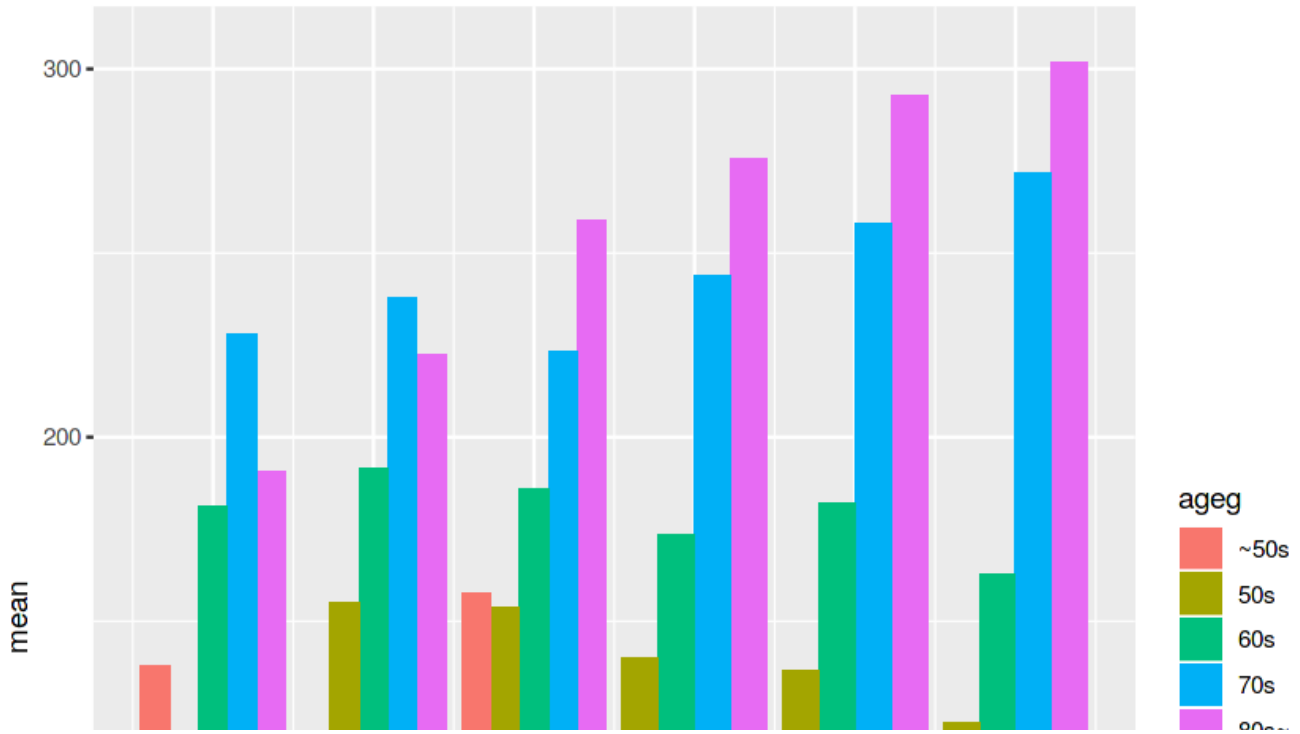
4. 경제적지원

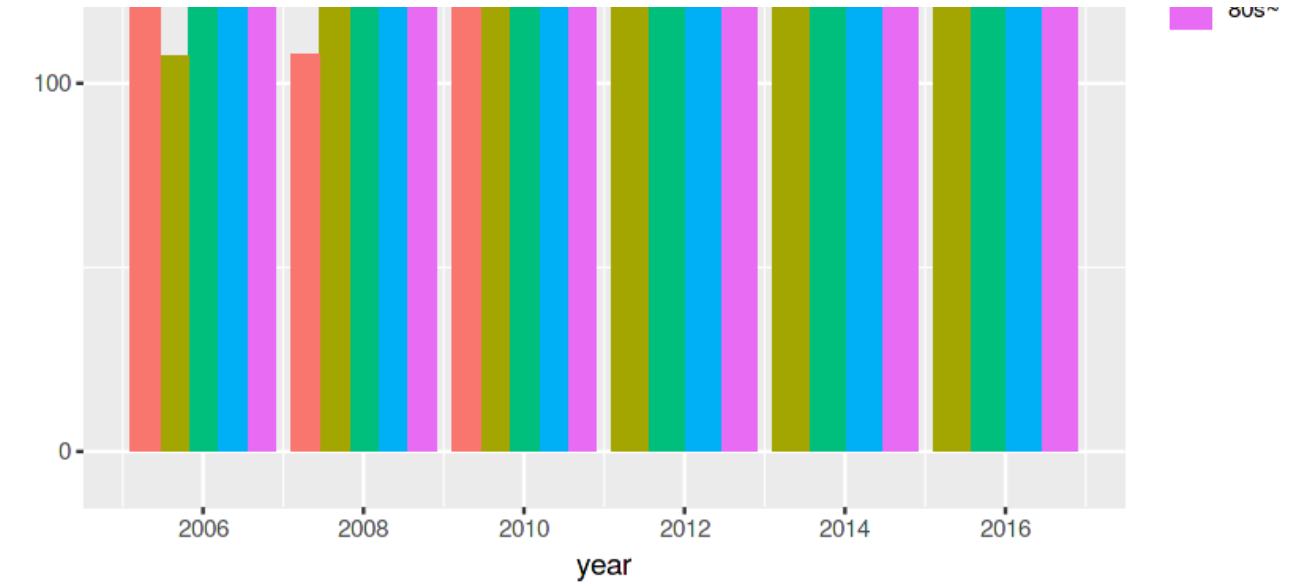
4.1 연도별 연령별 금전적 지원(자녀에게 지원 받은 금액)

In [41]:

```
klosa %>%
    filter(!is.na(m_fromchildren)) %>%
    select(year,ageg,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(mean = mean(m_fromchildren)) %>%
    ggplot(aes(year,mean,fill=ageg))+geom_boxplot() +
    scale_x_continuous(breaks=c(2006,2008,2010,2012,2014,2016))

# 이상치 제거 후 그래프그리기
klosa %>% filter(m_fromchildren<14000) -> klosa_from
klosa_from %>% filter(!is.na(m_fromchildren)) %>%
    select(year,ageg,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(mean = mean(m_fromchildren)) %>%
    ggplot(aes(year,mean,fill=ageg))+geom_col(position="dodge") +
    scale_x_continuous(breaks=c(2006,2008,2010,2012,2014,2016))
```


350
300
250
200
150
100
mean
2008
2010
2012
2014
year
ageg
~50s
50s
60s
70s
80s~

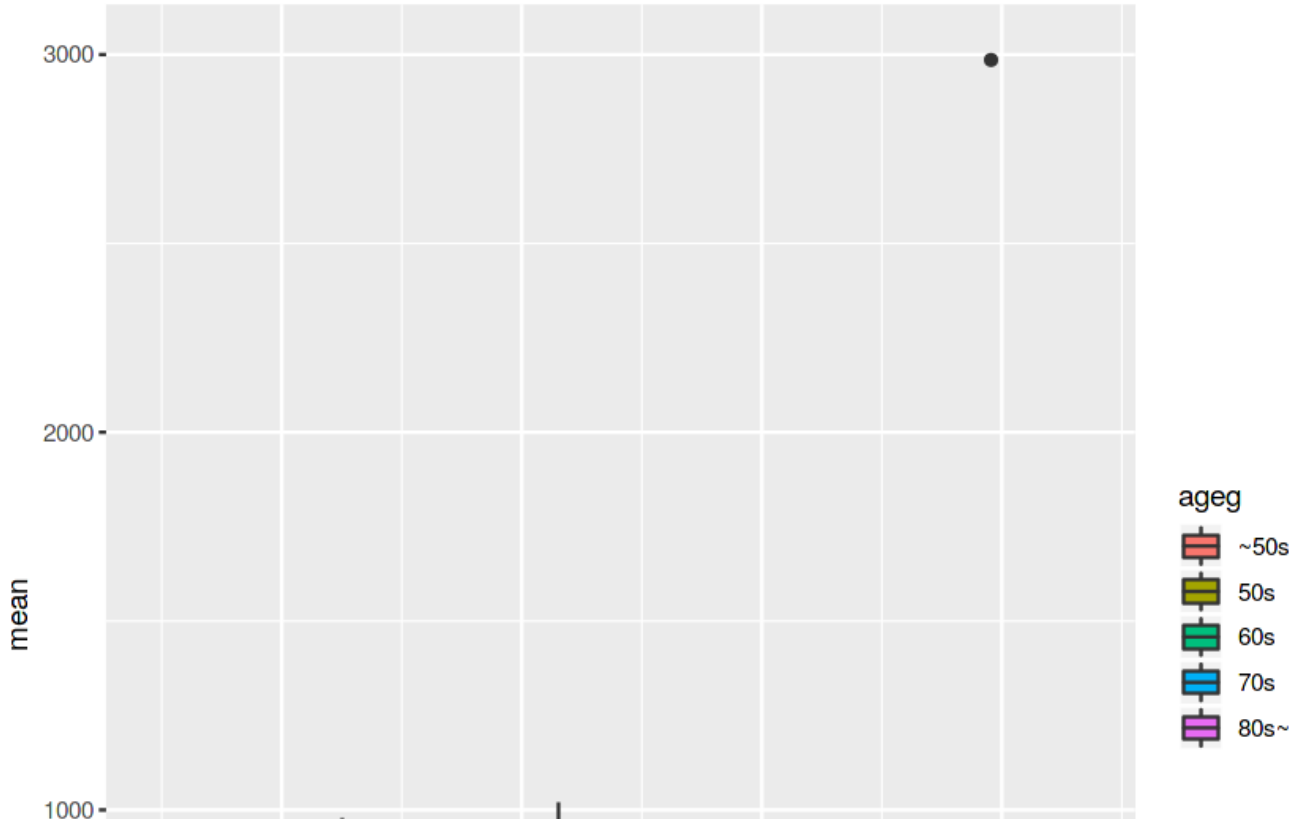

300
200
mean
ageg
~50s
50s
60s
70s

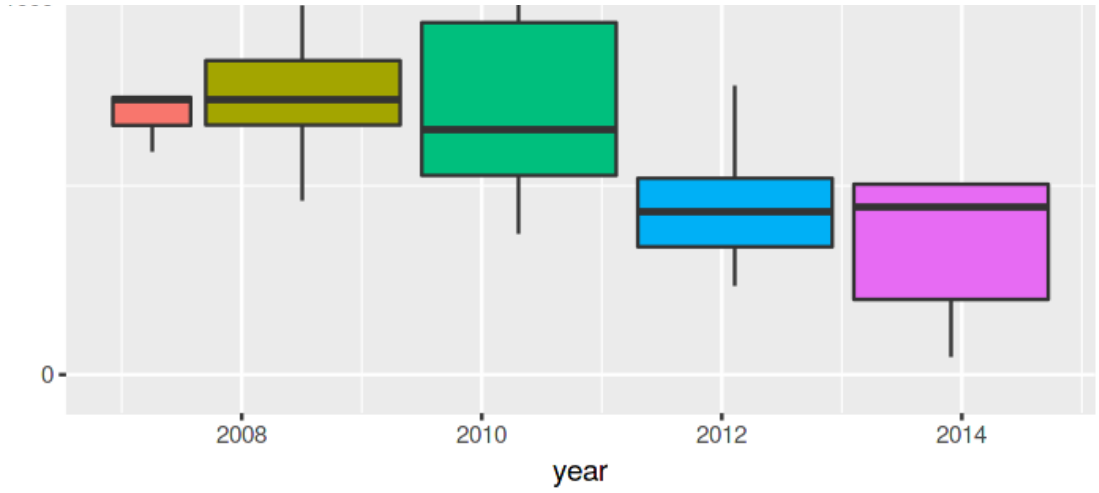


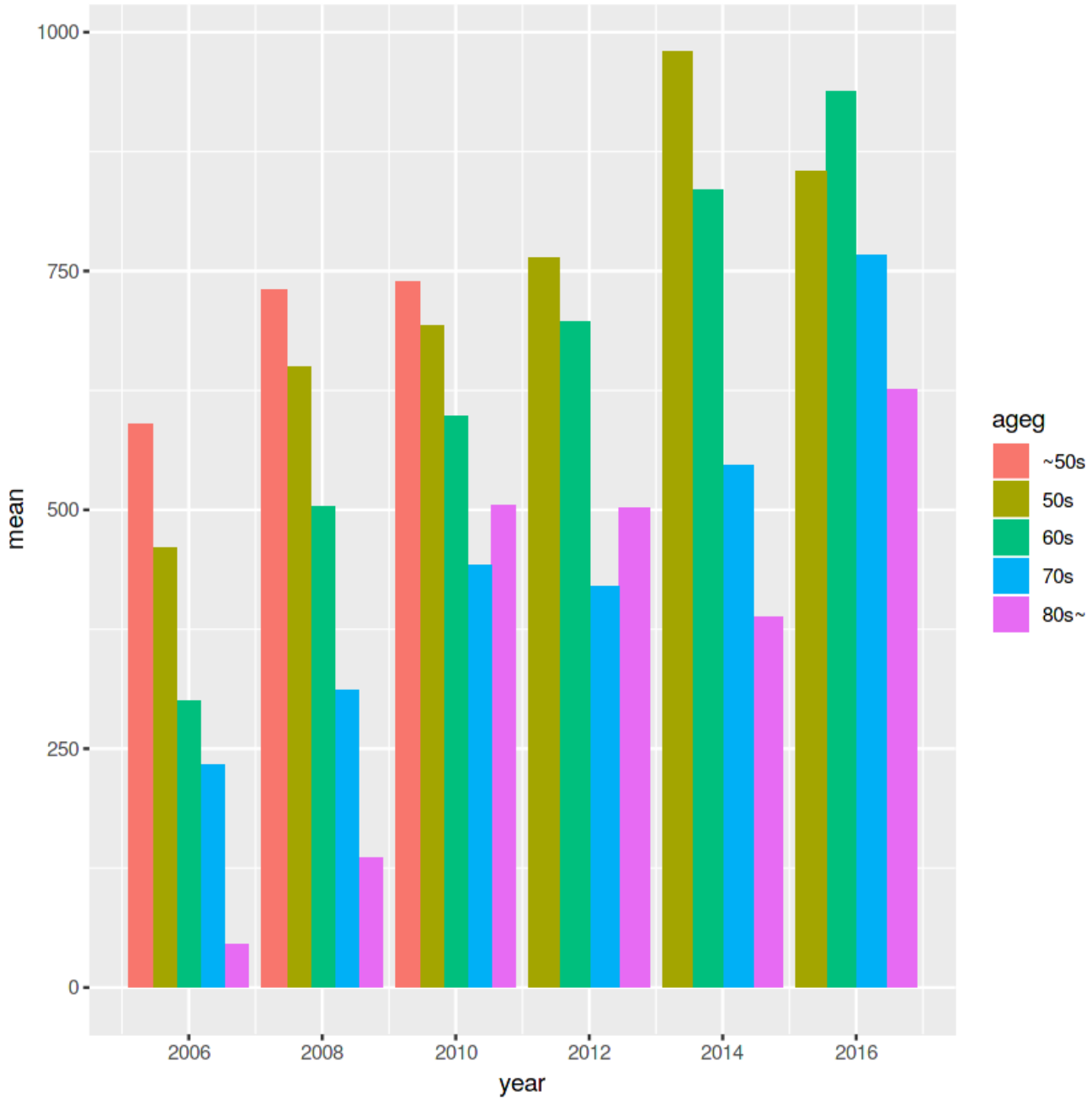
4.2 연도별 연령별 금전적 지원(자녀에게 지원한 금액)

In [42]:

```
klosa %>%
    filter(!is.na(m_tochildren)) %>%
    select(year,ageg,m_tochildren) %>%
    group_by(year,ageg) %>%
    summarise(mean = mean(m_tochildren)) %>%
    ggplot(aes(year,mean,fill=ageg))+geom_boxplot() +
    scale_x_continuous(breaks=c(2006,2008,2010,2012,2014,2016))

# 이상치 제거 후 그래프그리기
klosa %>% filter(m_tochildren<20000) -> klosa_to
klosa_to %>% filter(!is.na(m_tochildren)) %>%
    select(year,ageg,m_tochildren) %>%
    group_by(year,ageg) %>%
    summarise(mean = mean(m_tochildren)) %>%
    ggplot(aes(year,mean,fill=ageg))+geom_col(position="dodge") +
    scale_x_continuous(breaks=c(2006,2008,2010,2012,2014,2016))
```







4.3 연도별 금전적 지원

In [43]:

```
klosa %>%
```

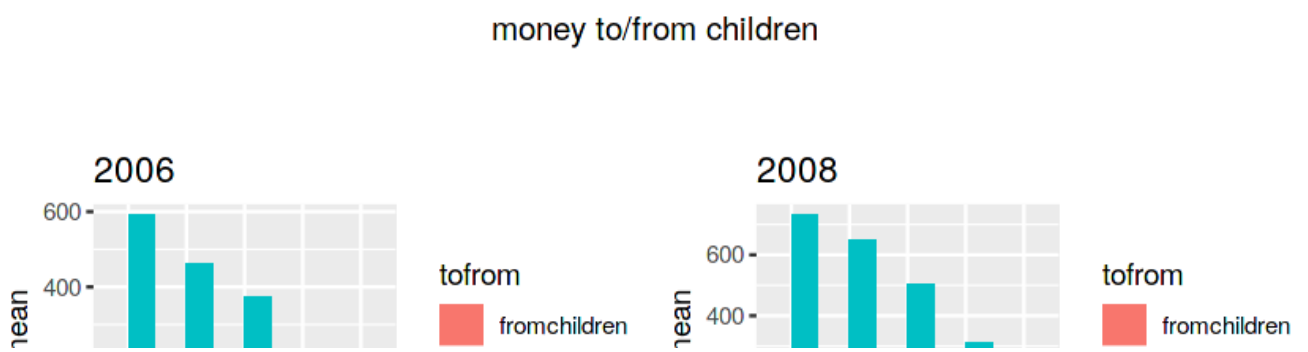
```
    filter(year==2006) %>%
    select(year,ageg,m_tochildren,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(tochildren= mean(m_tochildren, na.rm=T), fromchildren=mean(m_fromchildren, na.rm=T))
->money2006
klosa %>%
    filter(year==2008) %>%
    select(year,ageg,m_tochildren,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(tochildren=mean(m_tochildren, na.rm=T), fromchildren=mean(m_fromchildren, na.rm=T)) -
>money2008
klosa %>%
    filter(year==2010) %>%
    select(year,ageg,m_tochildren,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(tochildren=mean(m_tochildren, na.rm=T), fromchildren=mean(m_fromchildren, na.rm=T)) -
>money2010
klosa %>%
    filter(year==2012) %>%
    select(year,ageg,m_tochildren,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(tochildren=mean(m_tochildren, na.rm=T), fromchildren=mean(m_fromchildren, na.rm=T)) -
>money2012
klosa %>%
    filter(year==2014) %>%
    select(year,ageg,m_tochildren,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(tochildren=mean(m_tochildren, na.rm=T), fromchildren=mean(m_fromchildren, na.rm=T)) -
>money2014
klosa %>%
    filter(year==2016) %>%
    select(year,ageg,m_tochildren,m_fromchildren) %>%
    group_by(year,ageg) %>%
    summarise(tochildren=mean(m_tochildren, na.rm=T), fromchildren=mean(m_fromchildren, na.rm=T)) -
>money2016
```

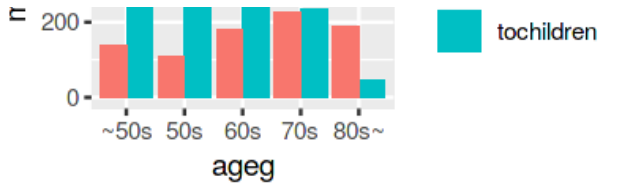
In [44]:

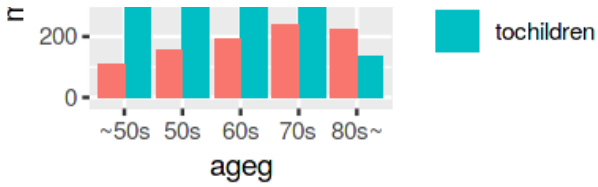
```
money_long2006=gather(money2006, key="tofrom", value="mean",`tochildren`,`fromchildren`)%>% ggplot(
aes(fill=tofrom,x=ageg,y=mean)) + geom_col(position="dodge")+
ggtitle('2006')
money_long2008=gather(money2008, key="tofrom", value="mean",`tochildren`,`fromchildren`)%>% ggplot(
aes(fill=tofrom,x=ageg,y=mean)) + geom_col(position="dodge")+
ggtitle('2008')
money_long2010=gather(money2010, key="tofrom", value="mean",`tochildren`,`fromchildren`)%>% ggplot(
aes(fill=tofrom,x=ageg,y=mean)) + geom_col(position="dodge")+
ggtitle('2010')
money_long2012=gather(money2012, key="tofrom", value="mean",`tochildren`,`fromchildren`)%>% ggplot(
aes(fill=tofrom,x=ageg,y=mean)) + geom_col(position="dodge")+
ggtitle('2012')
money_long2014=gather(money2014, key="tofrom", value="mean",`tochildren`,`fromchildren`)%>% ggplot(
aes(fill=tofrom,x=ageg,y=mean)) + geom_col(position="dodge")+
ggtitle('2014')
money_long2016=gather(money2016, key="tofrom", value="mean",`tochildren`,`fromchildren`)%>% ggplot(
aes(fill=tofrom,x=ageg,y=mean)) + geom_col(position="dodge")+
ggtitle('2016')
```

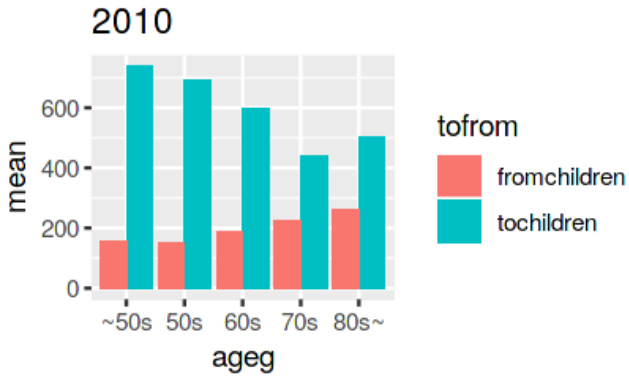
In [45]:

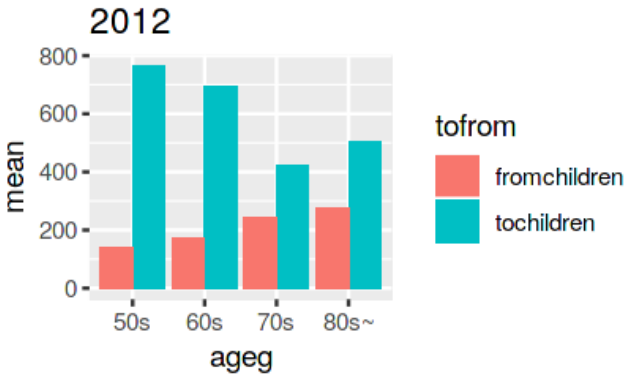
```
grid.arrange(money_long2006 ,money_long2008 ,money_long2010 ,money_long2012 ,money_long2014 ,money_
long2016 , ncol=2, top="money to/from children\n\n")
```

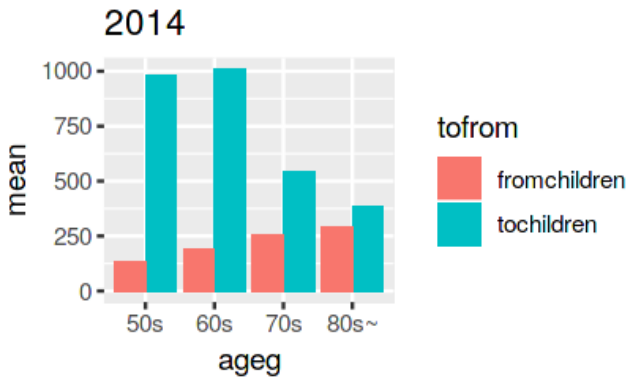


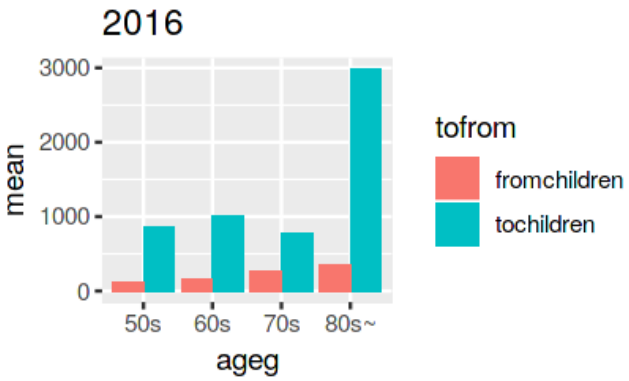












5. 소비

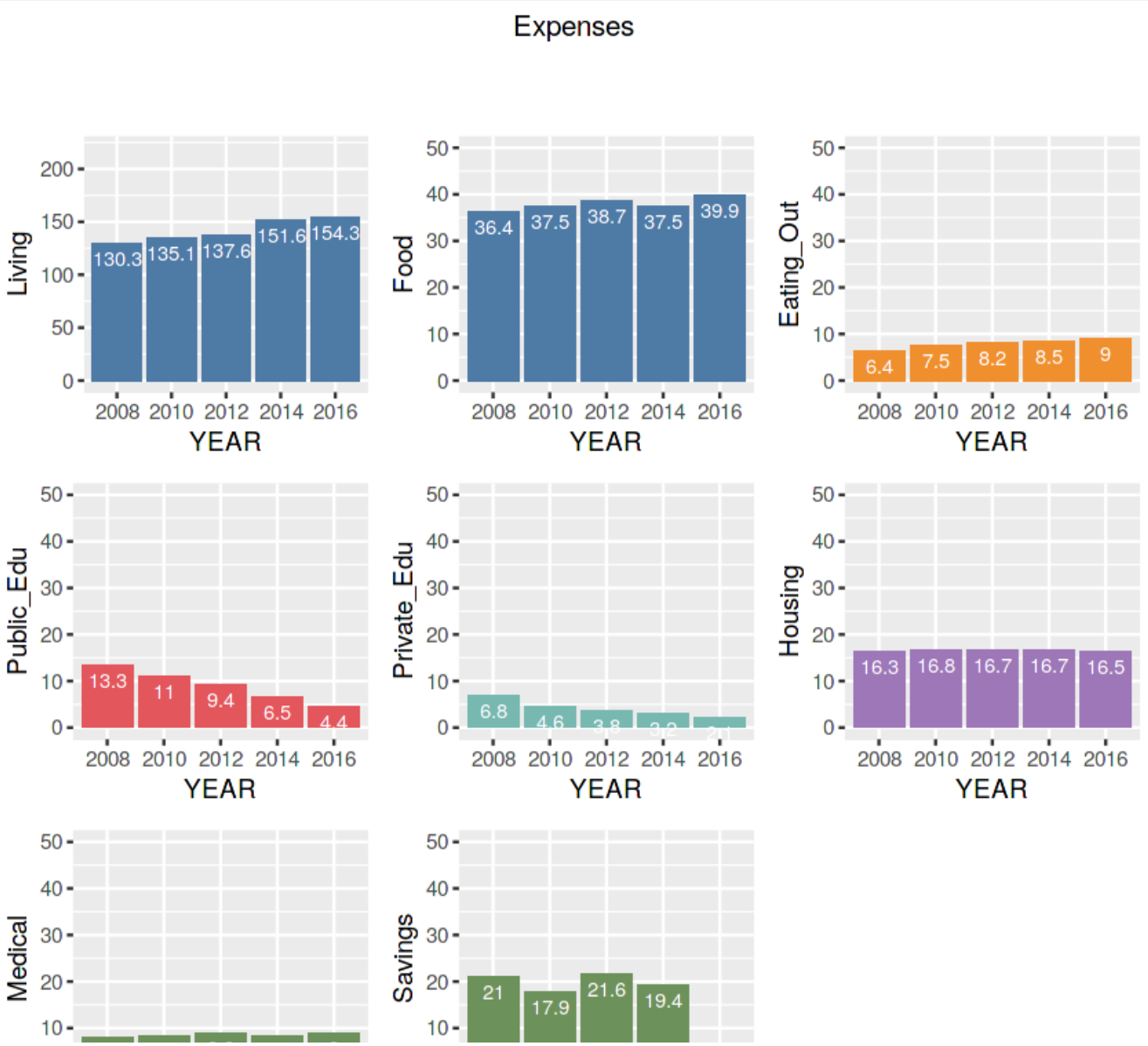
5.1 연도별 소비항목

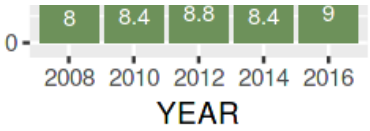
- 월평균 금액(단위:만원)
- living_expenses: 생활비
- food_expenses: 식비
- eating_out_expenses: 외식비
- public_education_expenses: 공교육비
- private_education_expenses: 사교육비
- housing_expenses: 주거비
- medical_expenses: 지보건의료비
- savings: 저축액(일반저축, 적금, 보험, 개인연금 등 포함)

In [46]:

```
klosa %>%
filter(year!=2006) %>%
select(year,living_expenses, food_expenses
,eating_out_expenses,public_education_expenses,private_education_expenses,housing_expenses,medical_
expenses,savings) %>%
group_by(year) %>%
summarize(living_expenses = sprintf("%.1f", mean(living_expenses,na.rm=T)),
          food_expenses=sprintf("%.1f", mean(food_expenses,na.rm=T)),
          eating_out_expenses=sprintf("%.1f", mean(eating_out_expenses,na.rm=T)),
          public_education_expenses=sprintf("%.1f", mean(public_education_expenses,na.rm=T)),
          private_education_expenses=sprintf("%.1f", mean(private_education_expenses,na.rm=T)),
          housing_expenses=sprintf("%.1f", mean(housing_expenses,na.rm=T)),
          medical_expenses=sprintf("%.1f", mean(medical_expenses,na.rm=T)),
          savings=sprintf("%.1f", mean(savings,na.rm=T))) -> expenses
expenses$living_expenses=as.numeric(expenses$living_expenses)
expenses$food_expenses=as.numeric(expenses$food_expenses)
expenses$eating_out_expenses=as.numeric(expenses$eating_out_expenses)
```

```
expenses$public_education_expenses=as.numeric(expenses$public_education_expenses)
expenses$private_education_expenses=as.numeric(expenses$private_education_expenses)
expenses$housing_expenses=as.numeric(expenses$housing_expenses)
expenses$medical_expenses=as.numeric(expenses$medical_expenses)
expenses$savings=as.numeric(expenses$savings)


expenses  %>% ggplot(aes(factor(year),living_expenses))+geom_col(fill="#4e7aa6")+
                                    geom_text(aes(label=living_expenses), vjust=1.5, colour="white"
size=3)+labs(x="YEAR",y="Living")+ylim(0,220)-> ex0
expenses  %>% ggplot(aes(factor(year),food_expenses))+geom_col(fill="#4e7aa6")+
                                    geom_text(aes(label=food_expenses), vjust=1.5, colour="white",
ize=3)+labs(x="YEAR",y="Food")+ylim(0,50)-> ex1
expenses  %>% ggplot(aes(factor(year),eating_out_expenses))+geom_col(fill="#f08f2e")+
                                    geom_text(aes(label=eating_out_expenses), vjust=1.5, colour="wh
te", size=3)+labs(x="YEAR",y="Eating_Out")+ylim(0,50)-> ex2
expenses  %>% ggplot(aes(factor(year),public_education_expenses))+geom_col(fill="#e3565b")+
                                    geom_text(aes(label=public_education_expenses), vjust=1.5, colc
r="white", size=3)+labs(x="YEAR",y="Public_Edu")+ylim(0,50)-> ex3
expenses  %>% ggplot(aes(factor(year),private_education_expenses))+geom_col(fill="#77b8b0")+
                                    geom_text(aes(label=private_education_expenses), vjust=1.5, col
ur="white", size=3)+labs(x="YEAR",y="Private_Edu")+ylim(0,50)-> ex4
expenses  %>% ggplot(aes(factor(year),housing_expenses))+geom_col(fill="#9d77b8")+
                                    geom_text(aes(label=housing_expenses), vjust=1.5, colour="white
, size=3)+labs(x="YEAR",y="Housing")+ylim(0,50)-> ex5
expenses  %>% ggplot(aes(factor(year),medical_expenses))+geom_col(fill="#6d915a")+
                                    geom_text(aes(label=medical_expenses), vjust=1.5, colour="white
, size=3)+labs(x="YEAR",y="Medical")+ylim(0,50)-> ex6
expenses  %>% ggplot(aes(factor(year),savings))+geom_col(fill="#6d915a")+
                                    geom_text(aes(label=savings), vjust=1.5, colour="white", size=3
+labs(x="YEAR",y="Savings")+ylim(0,50)-> ex7
grid.arrange(ex0,ex1,ex2,ex3,ex4,ex5,ex6,ex7 ,   nrow=3, top="Expenses\n\n")
```





6. 건강상태

6.1 ADL(Activities of Daily Living) - 일상생활 수행 능력

- change_clothes: 옷갈아입기
- bath: 목욕/샤워하기
- meal: 차려놓은 음식 식사하기
- out_of_room: 이부자리에서 일어나 방 밖으로 나오기
- toilet: 화장실 이용하기
- urin: 대소변 조절하기

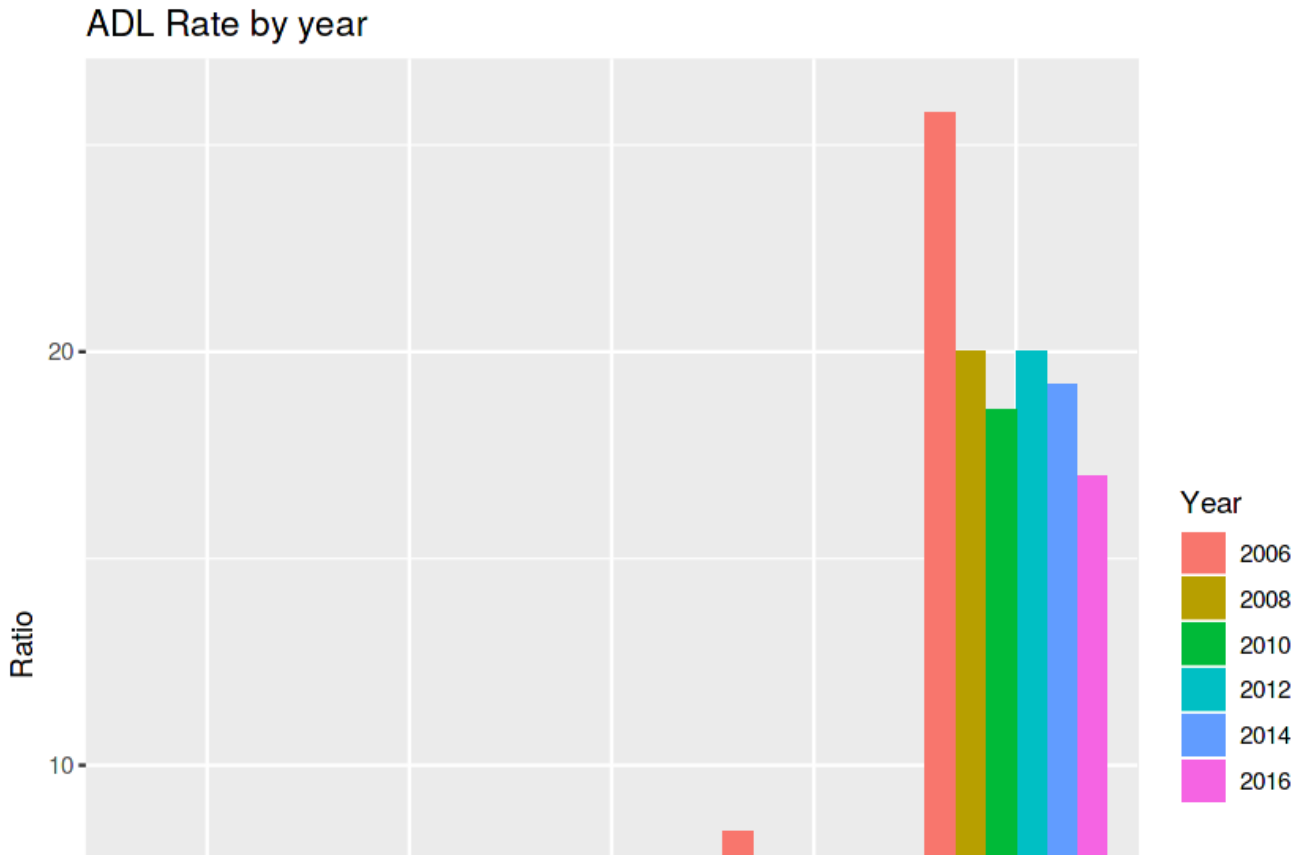
1) ADL 장애율

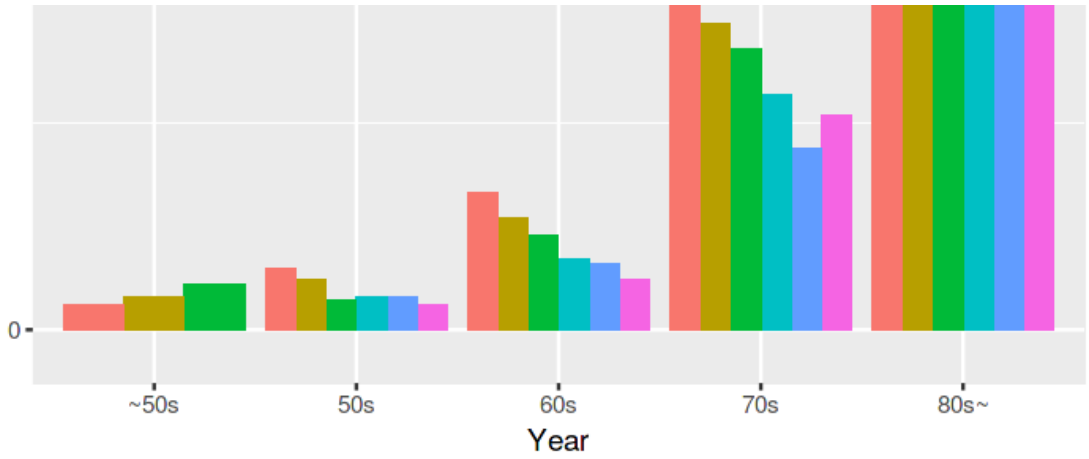
In [47]:

```
klosa %>% select(year,ageg, adl) %>%
    mutate(adl=ifelse(adl > 0, 1, 0)) %>%
    count(year,ageg,adl) %>%
    group_by(year,ageg) %>%
    mutate(total=sum(n)) %>%
    filter(adl==1) %>%
    mutate(pct=round(n/total*100,1))-> adl_rate
```

In [48]:

```
adl_rate %>% ggplot(aes(ageg,pct,fill=factor(year)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="ADL Rate by year") +
    theme_grey()+
    scale_fill_discrete(name = "Year")
```





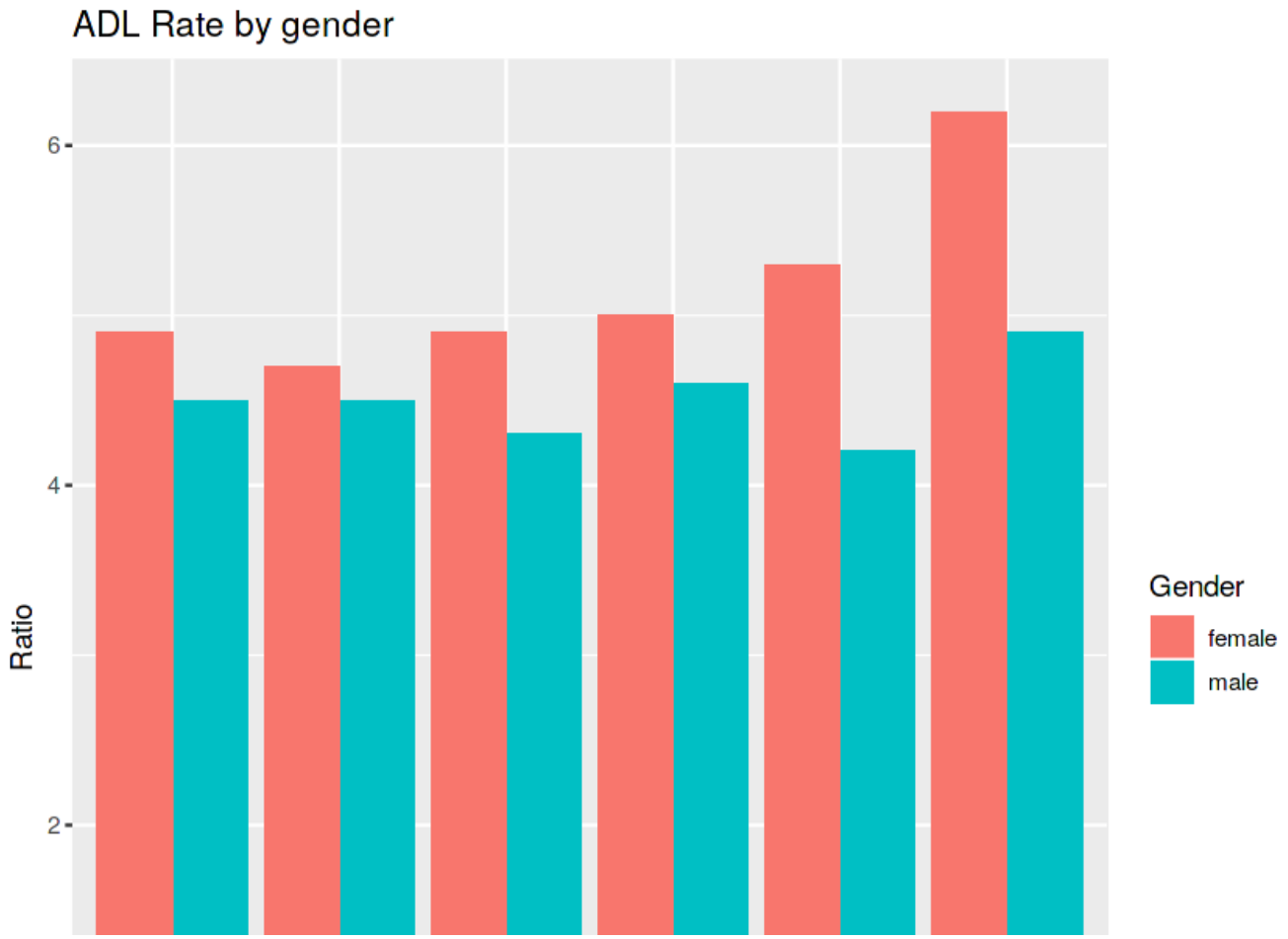
2) 성별 ADL 장애율

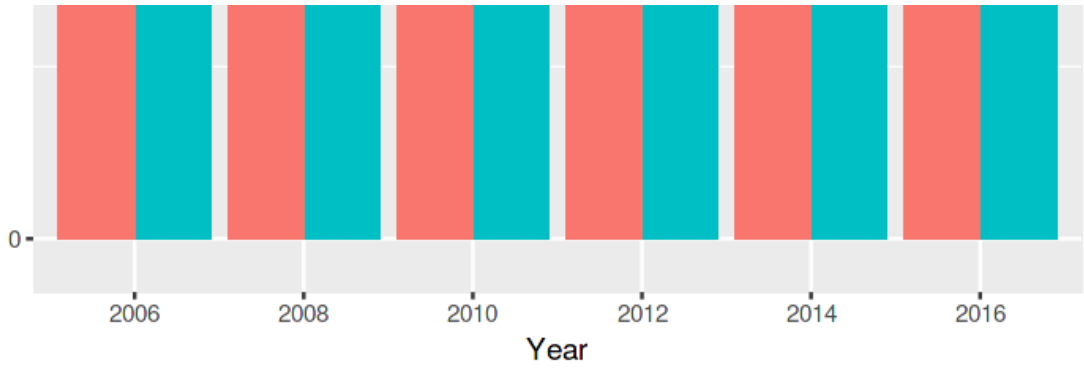
In [49]:

```
klosa %>% select(year,gender, adl) %>%
    mutate(adl=ifelse(adl > 0, 1, 0)) %>%
    count(year,gender,adl) %>%
    group_by(year,gender) %>%
    mutate(total=sum(n)) %>%
    filter(adl==1) %>%
    mutate(pct=round(n/total*100,1))-> adl_genrate
```

In [50]:

```
adl_genrate %>% ggplot(aes(factor(year),pct,fill=factor(gender)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="ADL Rate by gender") +
    theme_grey()+
    scale_fill_discrete(name = "Gender")
```





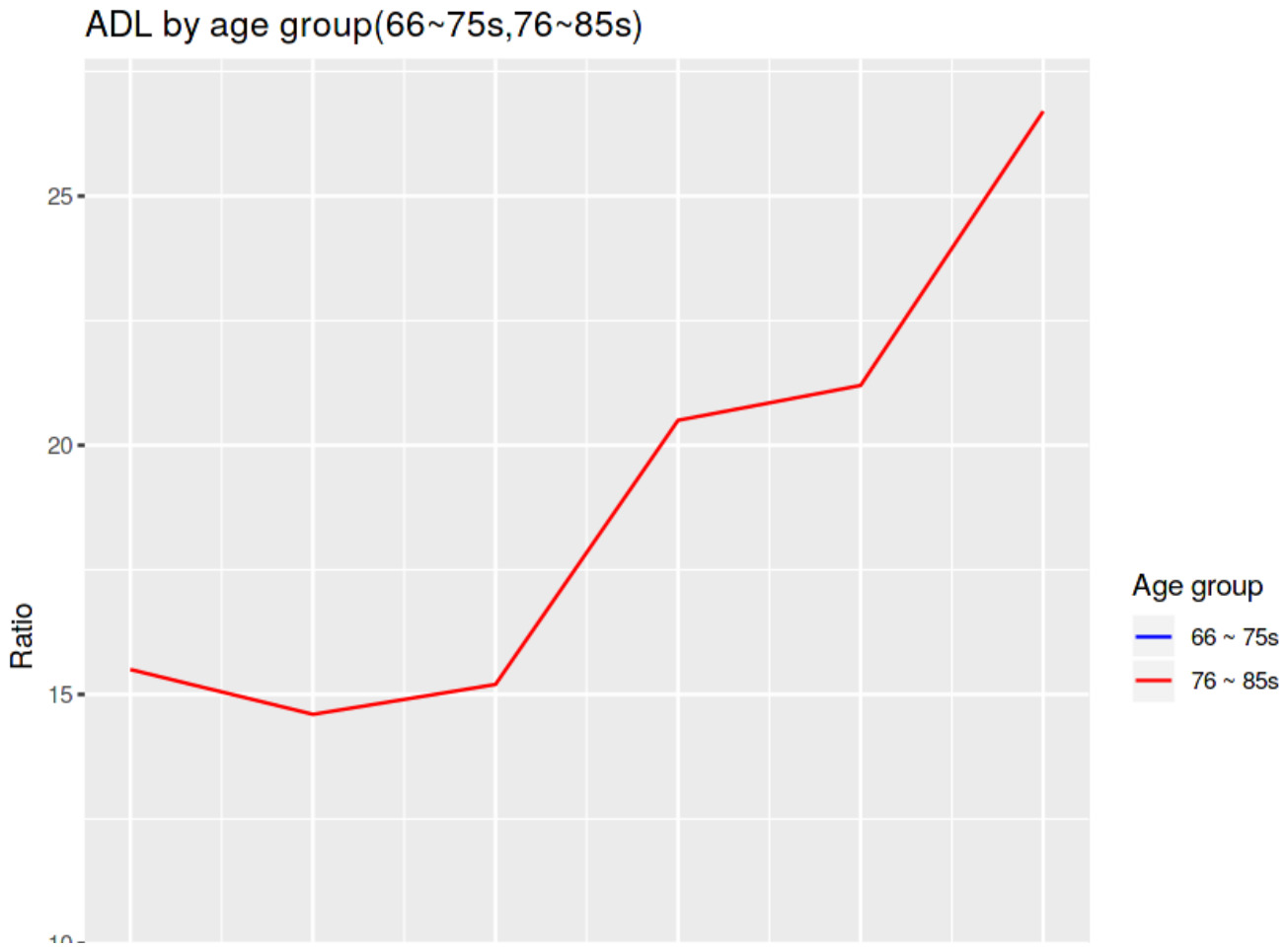
3) ADL 현황 ((2006년 66세 ~ 85세)

In [51]:

```
klosabyage %>% filter(!is.na(adl)&!is.na(year)) %>%
    select(year,age,tfr,adl) %>%
    mutate(adl=ifelse(adl > 0 , 1, 0)) %>%
    #filter(adl!=0) %>%
    count(year,tfr,adl) %>%
    group_by(year,tfr) %>%
    mutate(total=sum(n)) %>%
    filter(adl==1) %>%
    mutate(pct=round(n/total*100,1))->adl_line
```

In [52]:

```
ggplot(adl_line[adl_line$tfr %in% c("66to75","76to85"),],aes(year,pct,color=factor(tfr)))+
    geom_line()+
    labs(y="Ratio",x="Year",title="ADL by age group(66~75s,76~85s)")+
    scale_color_manual( values=c("Blue","Red"),
                        name = ("Age group"),
                        labels = c("66 ~ 75s","76 ~ 85s"))
```



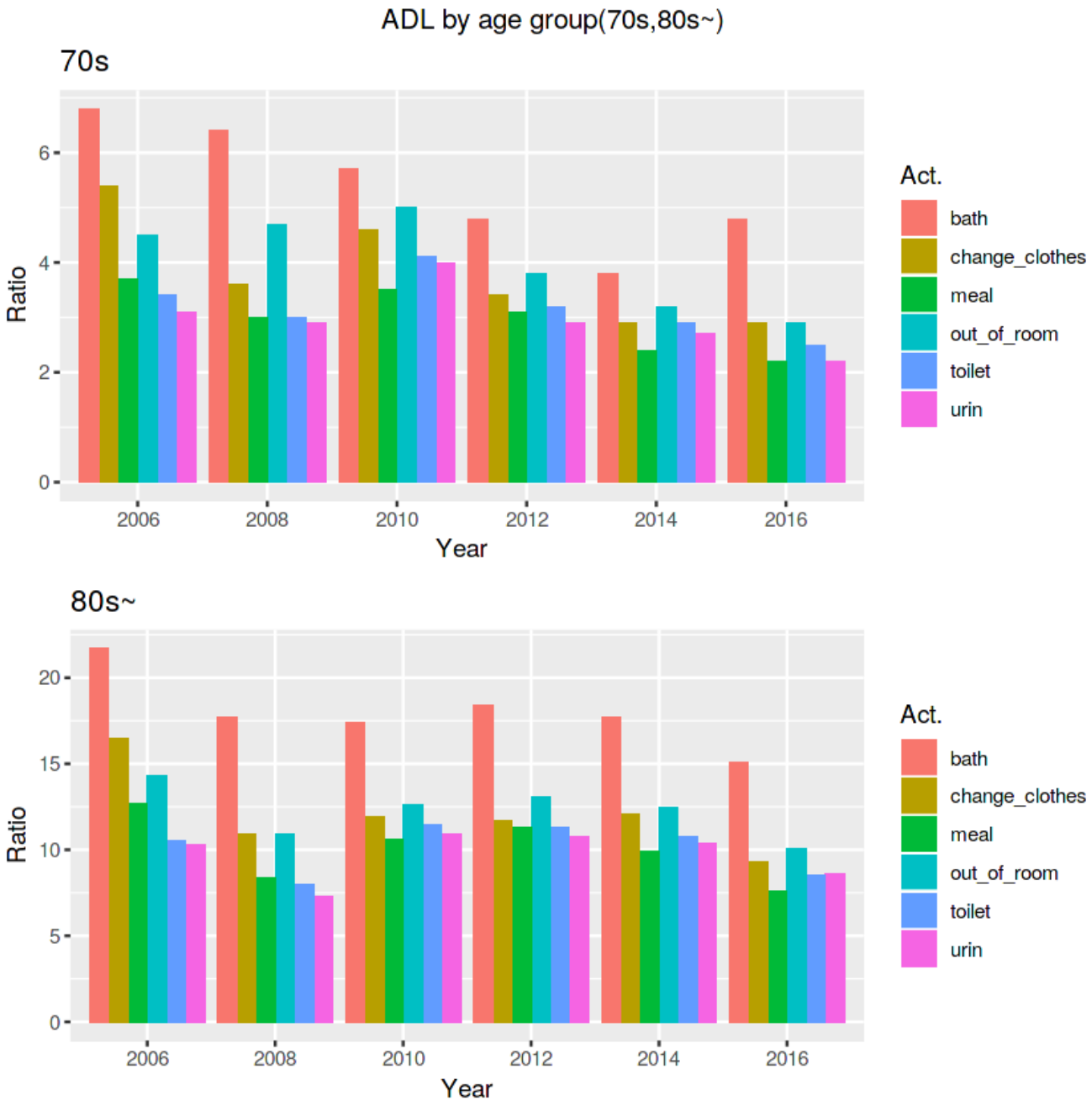
4) 일상생활 수행 장애율(70s, 80S~)

In [53]:

```
klosa %>% select(year,ageg,change_clothes,bath,meal,out_of_room,toilet,urin)  %>%
    filter(ageg=="70s"| ageg=="80s~") %>%
    mutate(change_clothes=ifelse(change_clothes==1,0,1)) %>%
    mutate(bath=ifelse(bath==1,0,1)) %>%
    mutate(meal=ifelse(meal==1,0,1)) %>%
    mutate(out_of_room=ifelse(out_of_room==1,0,1)) %>%
    mutate(toilet=ifelse(toilet==1,0,1)) %>%
    mutate(urin=ifelse(urin==1,0,1)) ->adl_count

adl_count %>% group_by(year,ageg,change_clothes)  %>%
    summarize(count=n()) %>%
    mutate(class="change_clothes") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(change_clothes!=0) ->m1

adl_count %>% group_by(year,ageg,bath)  %>%
    summarize(count=n()) %>%
    mutate(class="bath") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(bath!=0)->m2

adl_count %>% group_by(year,ageg,meal)  %>%
    summarize(count=n()) %>%
    mutate(class="meal") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(meal!=0)->m3

adl_count %>% group_by(year,ageg,out_of_room)  %>%
    summarize(count=n()) %>%
    mutate(class="out_of_room") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(out_of_room!=0)->m4

adl_count %>% group_by(year,ageg,toilet)  %>%
    summarize(count=n()) %>%
    mutate(class="toilet") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(toilet!=0)->m5

adl_count %>% group_by(year,ageg,urin)  %>%
    summarize(count=n()) %>%
    mutate(class="urin") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(urin!=0)->m6

rbind(m1,m2,m3,m4,m5,m6) -> adl_bind
adl_bind %>%  select(year,ageg,class,count,total,pct)->adl_bind
```

In [54]:

```
ggplot(adl_bind[adl_bind$ageg=="70s",],aes(factor(year),pct,fill=factor(class)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="70s") +
    theme_grey()+
    scale_fill_discrete(name = "Act.")->adl1

ggplot(adl_bind[adl_bind$ageg=="80s~",],aes(factor(year),pct,fill=factor(class)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="80s~") +
    theme_grey()+
    scale_fill_discrete(name = "Act.")->adl2

grid.arrange(adl1, adl2, top="ADL by age group(70s,80s~)", nrow = 2)
```



6.2 IADL(Instrumental Activities of Daily Living) - 도구적 일상 활동 수행 능력

- prepare_meal: 식사준비
- gloom: 몸단장 하기
- clean: 일상적인 집안일(청소 등)
- laun: 빨래하기(세탁, 빨래 널고 말리기 등)
- out: 가까운 거리 외출(교통수단 미이용)
- bus: 교통수단 이용하여 외출하기

- buying: 상점에서 물건사기
- m_manage: 금전관리(용돈/통장/재산관리 등)
- phone_call: 전화걸고 받기
- medicine: 약 챙겨 먹기

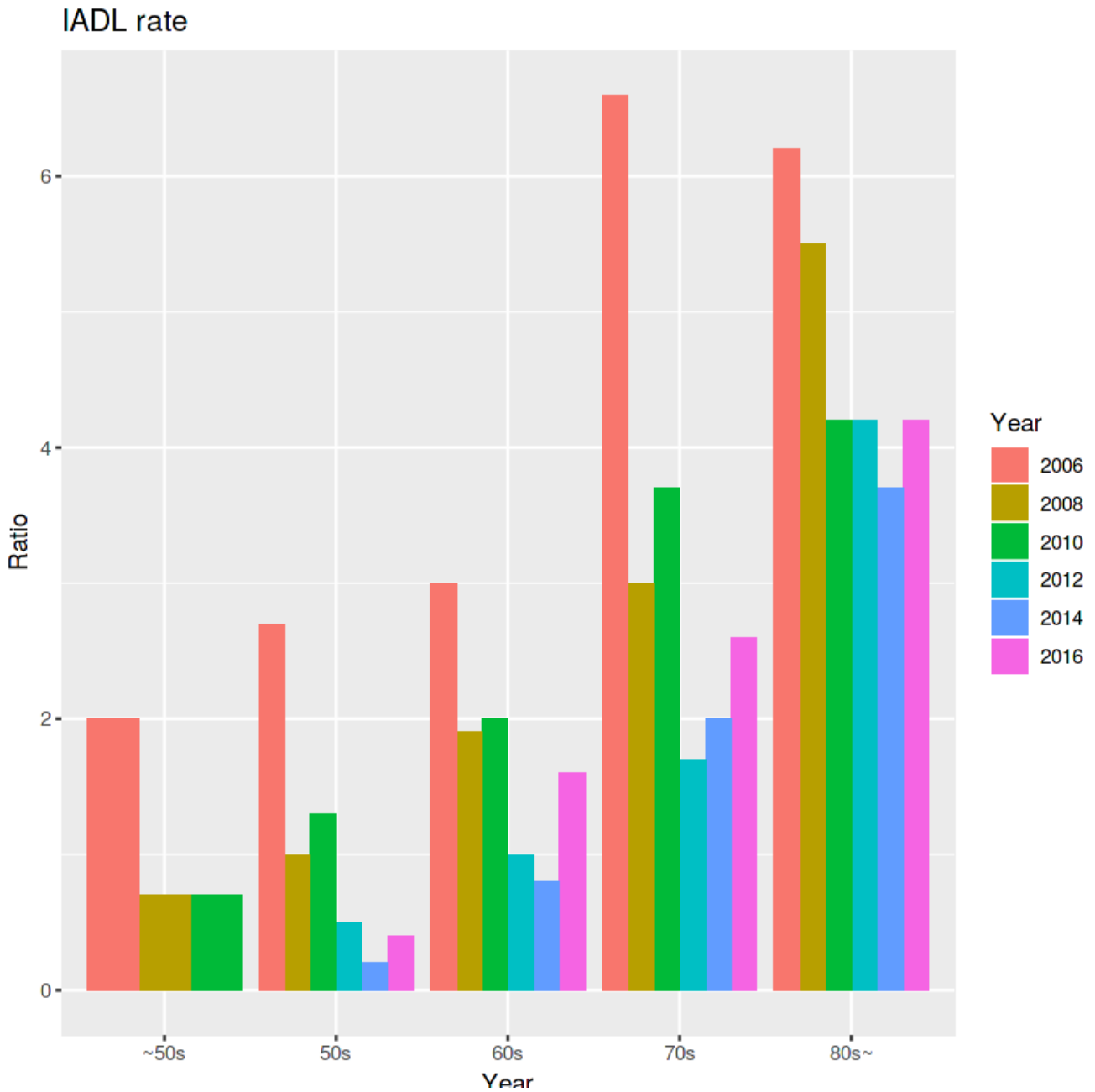
1) IADL 장애율

In [55]:

```
klosa %>% select(year,ageg, iadl) %>%
    mutate(adl=ifelse(iadl > 0, 1, 0)) %>%
    count(year,ageg,iadl) %>%
    group_by(year,ageg) %>%
    mutate(total=sum(n)) %>%
    filter(iadl==1) %>%
    mutate(pct=round(n/total*100,1))-> iadl_rate
```

In [56]:

```
iadl_rate %>% ggplot(aes(ageg,pct,fill=factor(year)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="IADL rate") +
    theme_grey()+
    scale_fill_discrete(name = "Year")
```



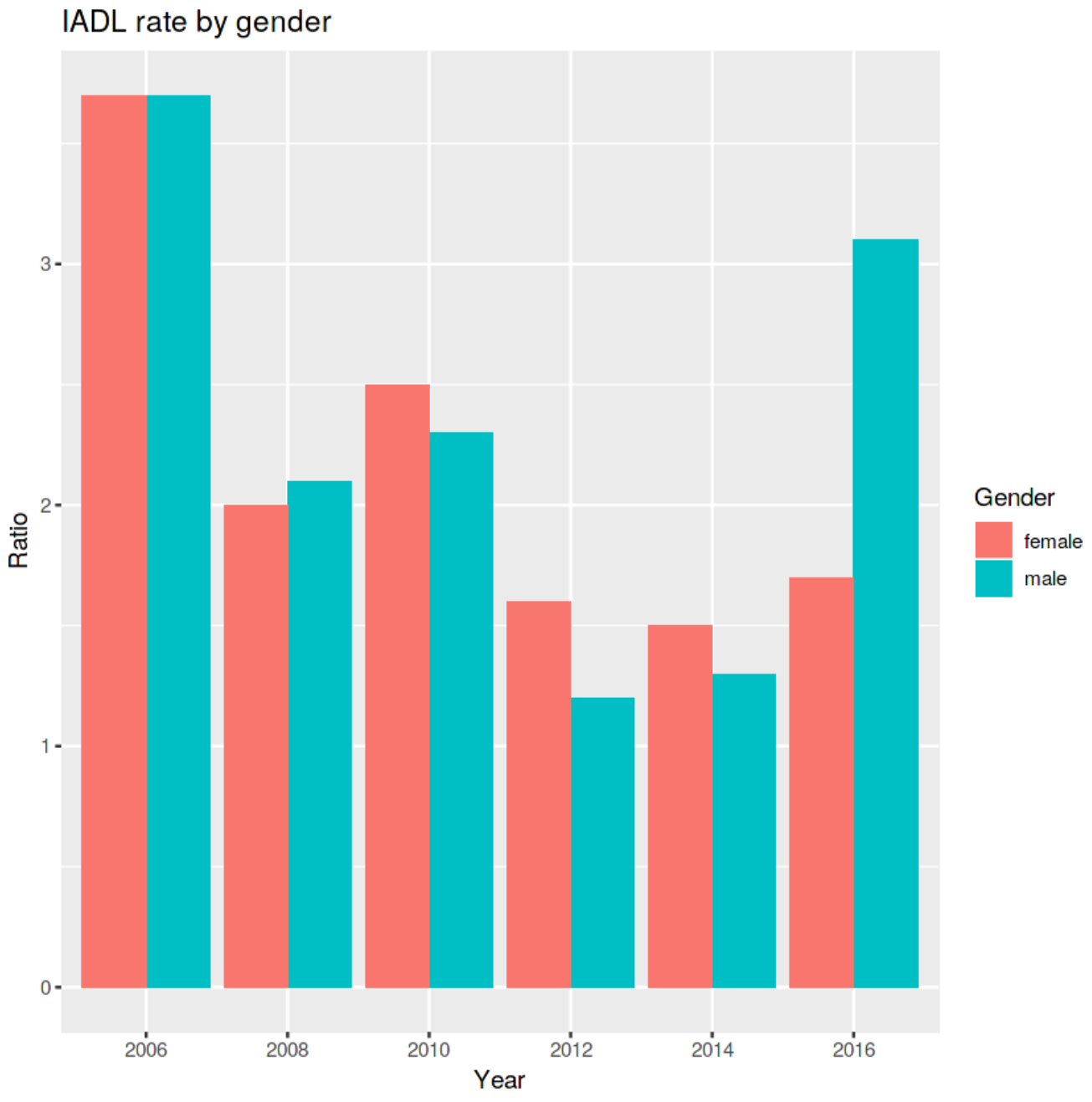
2) 성별 IADL 장애율

In [57]:

```
klosa %>% select(year,gender, iadl) %>%
    mutate(adl=ifelse(iadl > 0, 1, 0)) %>%
    count(year,gender,iadl) %>%
    group_by(year,gender) %>%
    mutate(total=sum(n)) %>%
    filter(iadl==1) %>%
    mutate(pct=round(n/total*100,1))-> iadl_genrate
```

In [58]:

```
iadl_genrate %>% ggplot(aes(factor(year),pct,fill=factor(gender)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="IADL rate by gender") +
    theme_grey()+
    scale_fill_discrete(name = "Gender")
```



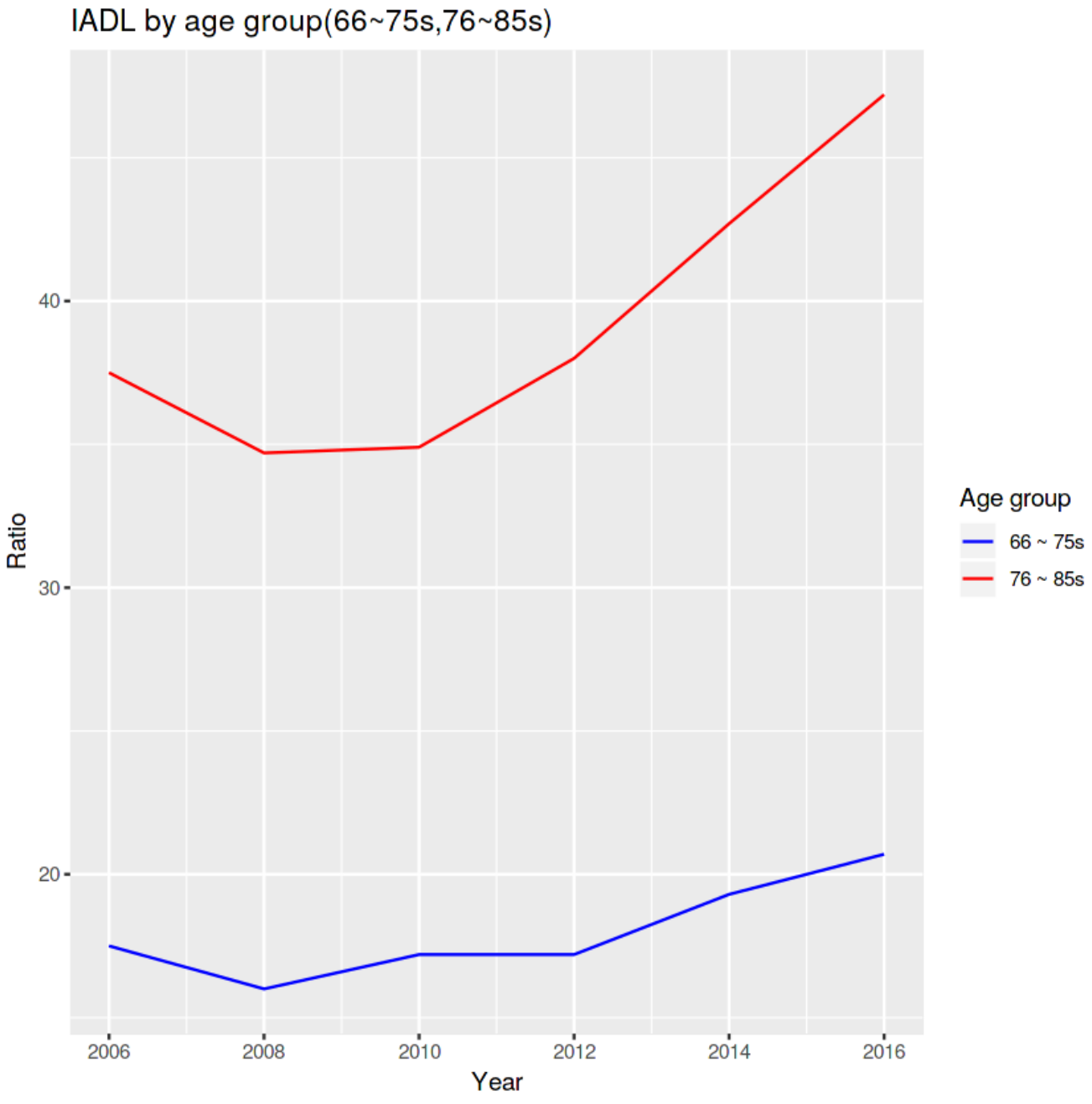
3) IADL 현황 ((2006년 66세 ~ 85세)

In [59]:

```
klosabyage %>% filter(!is.na(iadl)&!is.na(year)) %>%
    select(year,age,tfr,iadl) %>%
    mutate(iadl=ifelse(iadl > 0, 1, 0)) %>%
    #filter(adl!=0) %>%
    count(year,tfr,iadl) %>%
    group_by(year,tfr) %>%
    mutate(total=sum(n)) %>%
    filter(iadl==1) %>%
    mutate(pct=round(n/total*100,1))->iadl_line
```

In [60]:

```
ggplot(iadl_line[iadl_line$tfr %in% c("66to75","76to85"),],aes(year,pct,color=factor(tfr)))+
    geom_line()+
    labs(y="Ratio",x="Year",title="IADL by age group(66~75s,76~85s)")+
    scale_color_manual( values=c("Blue","Red"),
                        name = ("Age group"),
                        labels = c("66 ~ 75s","76 ~ 85s"))
```



4) 도구적 일상 활동 장애율(70s, 80S~)

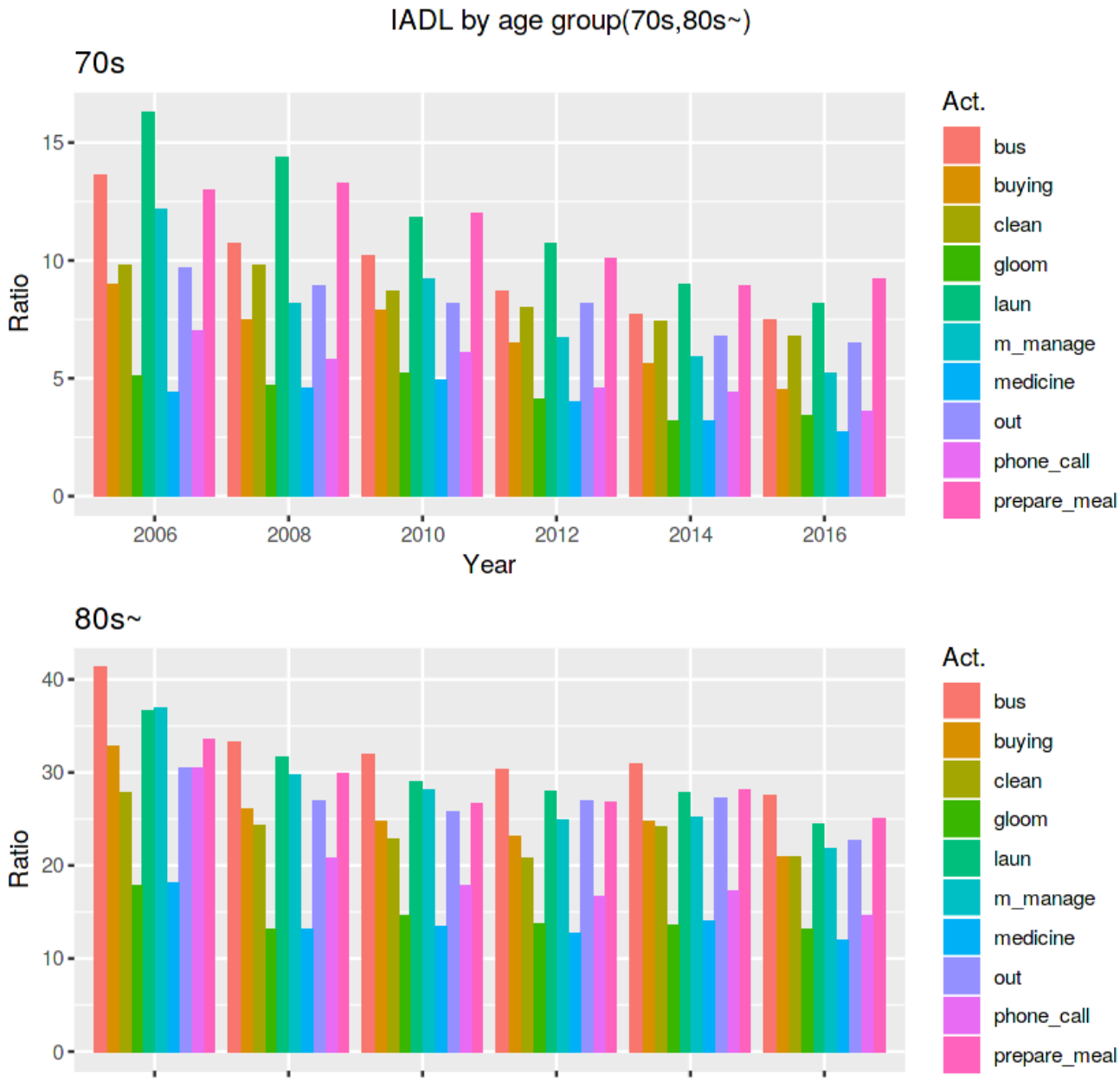
In [61]:

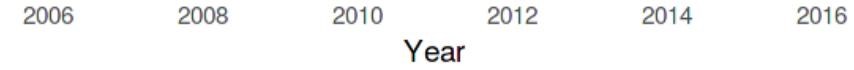
```
klosa %>% select(year,ageg,prepare_meal,gloom,clean,laun,out,bus,buying,m_manage,phone_call,medicin
e) %>%
          mutate(prepare_meal=ifelse(prepare_meal==1,0,1)) %>%
          mutate(gloom=ifelse(gloom==1,0,1)) %>%
          mutate(clean=ifelse(clean==1,0,1)) %>%
          mutate(laun=ifelse(laun==1,0,1)) %>%
          mutate(out=ifelse(out==1,0,1)) %>%
          mutate(bus=ifelse(bus==1,0,1)) %>%
          mutate(buying=ifelse(buying==1,0,1)) %>%
          mutate(m_manage=ifelse(m_manage==1,0,1)) %>%
          mutate(phone_call=ifelse(phone_call==1,0,1)) %>%
          mutate(medicine=ifelse(medicine==1,0,1)) ->iadl_count

iadl_count %>% group_by(year,ageg,prepare_meal)  %>%
    summarize(count=n()) %>%
    mutate(class="prepare_meal") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(prepare_meal!=0) ->n1

iadl_count %>% group_by(year,ageg,gloom)  %>%
    summarize(count=n()) %>%
    mutate(class="gloom") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(gloom!=0) ->n2

iadl_count %>% group_by(year,ageg,clean)  %>%
    summarize(count=n()) %>%
    mutate(class="clean") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(clean!=0) ->n3

iadl_count %>% group_by(year,ageg,laun)  %>%
    summarize(count=n()) %>%
    mutate(class="laun") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(laun!=0) ->n4

iadl_count %>% group_by(year,ageg,out)  %>%
    summarize(count=n()) %>%
    mutate(class="out") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(out!=0) ->n5

iadl_count %>% group_by(year,ageg,bus)  %>%
    summarize(count=n()) %>%
    mutate(class="bus") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(bus!=0) ->n6

iadl_count %>% group_by(year,ageg,buying)  %>%
    summarize(count=n()) %>%
    mutate(class="buying") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(buying!=0) ->n7

iadl_count %>% group_by(year,ageg,m_manage)  %>%
    summarize(count=n()) %>%
    mutate(class="m_manage") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(m_manage!=0) ->n8

iadl_count %>% group_by(year,ageg,phone_call)  %>%
    summarize(count=n()) %>%
    mutate(class="phone_call") %>%
```

```
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(phone_call!=0) ->n9

iadl_count %>% group_by(year,ageg,medicine)  %>%
    summarize(count=n()) %>%
    mutate(class="medicine") %>%
    mutate(total=sum(count)) %>%
    mutate(pct=round(count/total*100,1)) %>%
    filter(medicine!=0) ->n10

rbind(n1,n2,n3,n4,n5,n6,n7,n8,n9,n10) -> iadl_bind
iadl_bind %>%  select(year,ageg,class,count,total,pct)->iadl_bind
```

In [62]:

```
ggplot(iadl_bind[iadl_bind$ageg=="70s",],aes(factor(year),pct,fill=factor(class)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="70s") +
    theme_grey()+
    scale_fill_discrete(name = "Act.")->iadl1

ggplot(iadl_bind[iadl_bind$ageg=="80s~",],aes(factor(year),pct,fill=factor(class)))+
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="80s~") +
    theme_grey()+
    scale_fill_discrete(name = "Act.")->iadl2

grid.arrange(iadl1, iadl2, top="IADL by age group(70s,80s~)", nrow = 2)
```





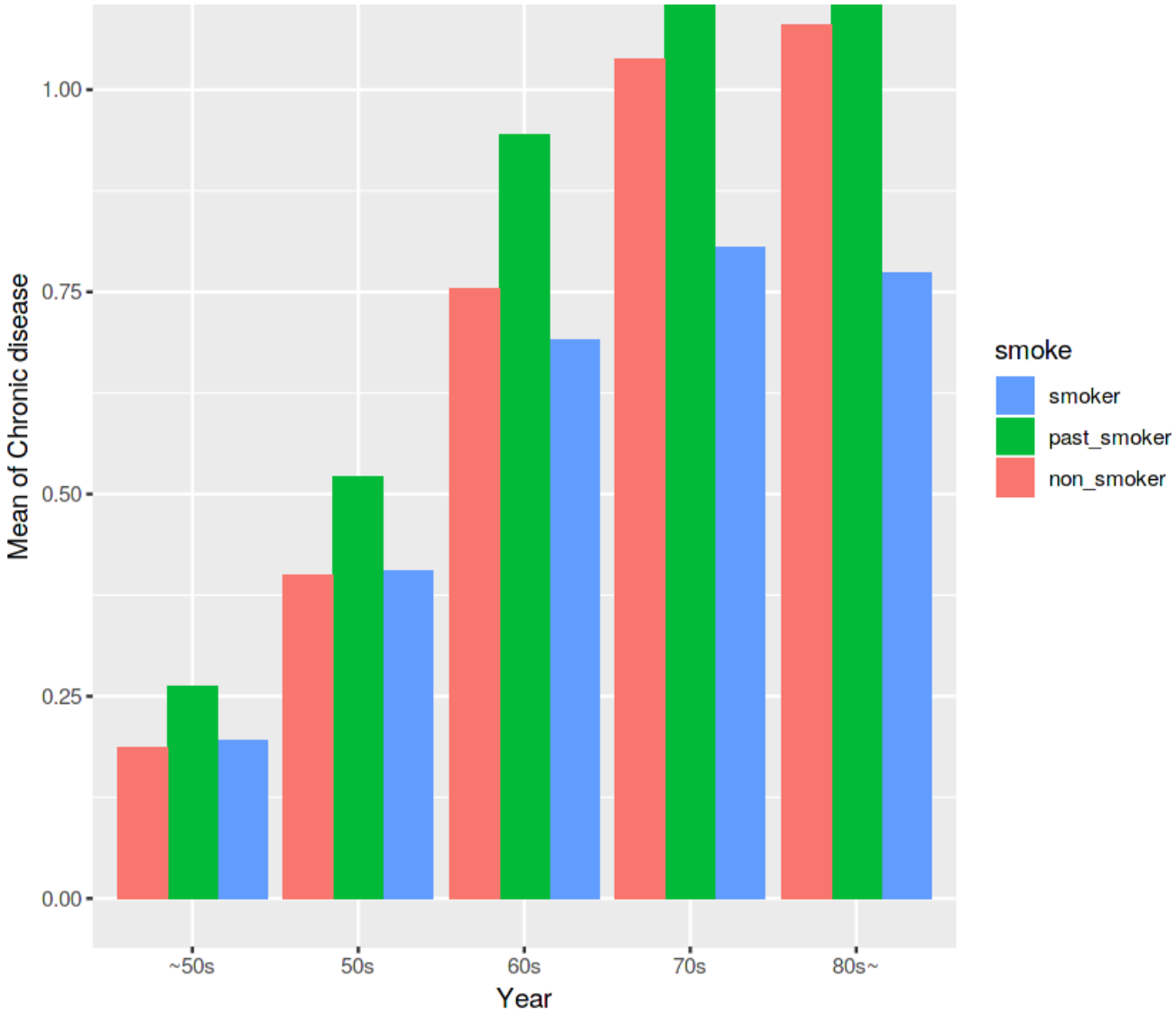
6.3 음주 및 흡연과 만성질환

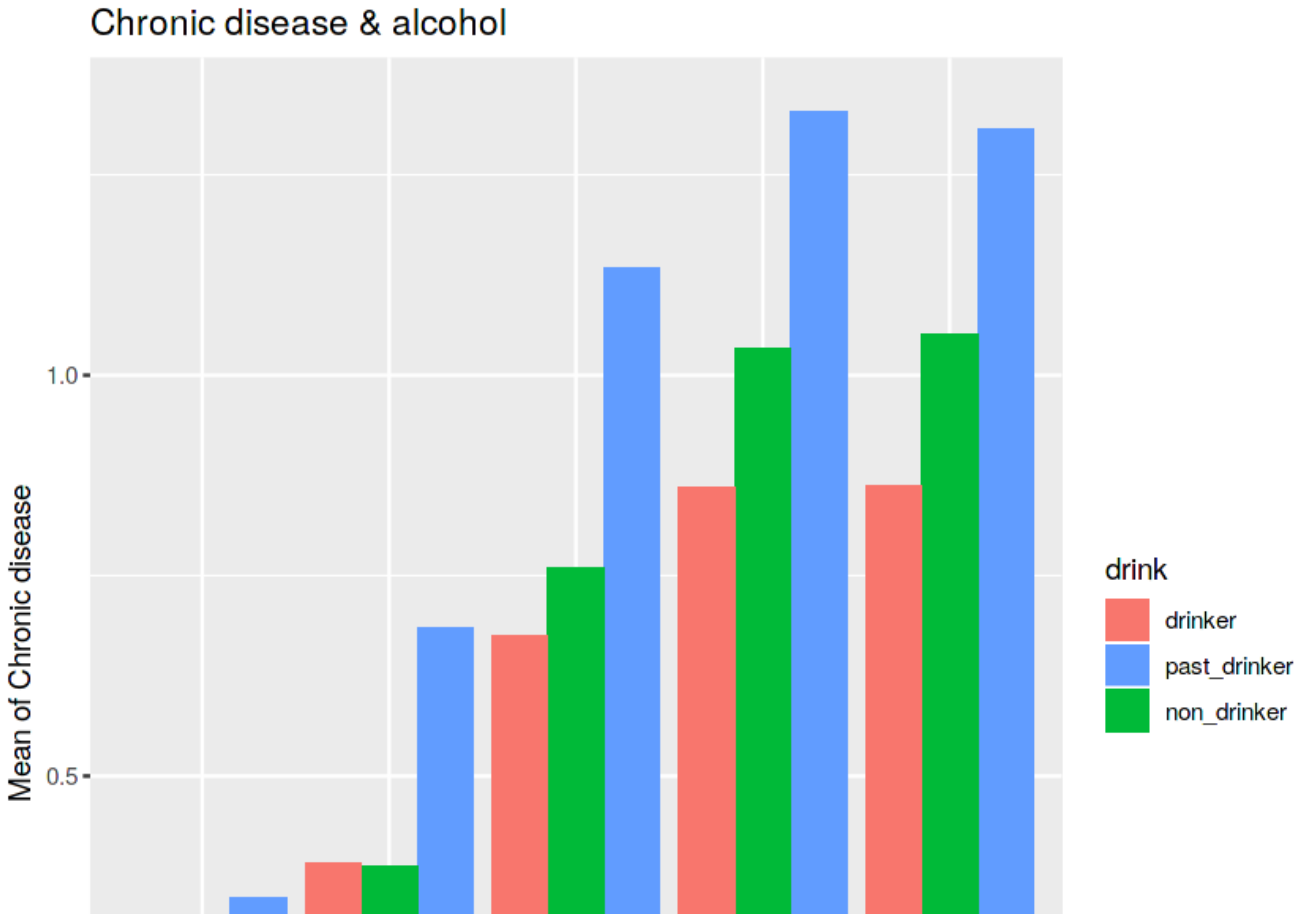
In [63]:

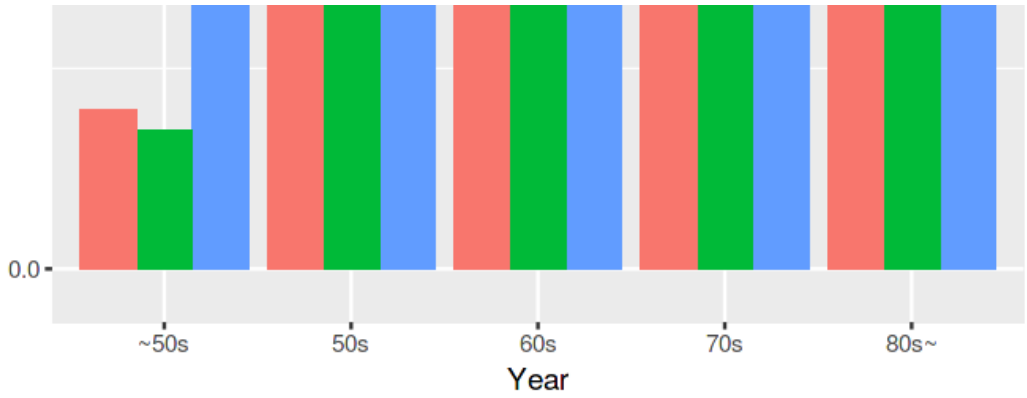
```
klosa %>%
select(year,gender,ageg,alcohol,smoke,hypertension,diabetes,cancer,lung_disease,liver_disease,heart
_disease,cerebrovascular_disease) %>%
    mutate(hypertension=ifelse(hypertension==1,1,0)) %>%
    mutate(diabetes=ifelse(diabetes==1,1,0)) %>%
    mutate(cancer=ifelse(cancer==1,1,0)) %>%
    mutate(lung_disease =ifelse(lung_disease==1,1,0)) %>%
    mutate(liver_disease=ifelse(liver_disease==1,1,0)) %>%
    mutate(heart_disease=ifelse(heart_disease==1,1,0)) %>%
    mutate(cerebrovascular_disease=ifelse(cerebrovascular_disease==1,1,0)) %>%
    mutate(chronic_sum=hypertension+diabetes+cancer+lung_disease+liver_disease+heart_disease+cerebr
ovascular_disease)->cerebrovascular_data
```

In [64]:

```
#흡연자와 평균 만성질환갯수
cerebrovascular_data %>% filter(!is.na(smoke)) %>% filter(!is.na(chronic_sum)) %>%
  filter(smoke==2) %>% select(ageg, chronic_sum) %>%
  group_by(ageg) %>% summarize(smoker=mean(chronic_sum)) ->smo1


cerebrovascular_data %>% filter(!is.na(smoke)) %>% filter(!is.na(chronic_sum)) %>%
  filter(smoke==1) %>% select(ageg, chronic_sum) %>%
  group_by(ageg) %>% summarize(past_smoker=mean(chronic_sum))->smo2

cerebrovascular_data %>% filter(!is.na(smoke)) %>% filter(!is.na(chronic_sum)) %>%
  filter(smoke==0) %>% select(ageg, chronic_sum) %>%
  group_by(ageg) %>% summarize(non_smoker=mean(chronic_sum))->smo3

left_join(smo1,smo2, by='ageg')->smoke
left_join(smoke,smo3, by='ageg')->smoke
gather(smoke, key="smoke", value="mean",`smoker`,`past_smoker`,`non_smoker`) %>%
  ggplot(aes(factor(ageg),mean,fill=smoke))+geom_col(position="dodge")+
  scale_fill_discrete(breaks=c("smoker","past_smoker","non_smoker"))+
  labs(y="Mean of Chronic disease",x="Year",title="Chronic disease & smok")

#음주자와 평균 만성질환갯수
cerebrovascular_data %>% filter(!is.na(alcohol)) %>% filter(!is.na(chronic_sum)) %>%
    filter(alcohol==1) %>% select(ageg, chronic_sum) %>%
    group_by(ageg) %>% summarize(drinker=mean(chronic_sum)) ->ac1


  cerebrovascular_data %>% filter(!is.na(alcohol)) %>% filter(!is.na(chronic_sum)) %>%
    filter(alcohol==2) %>% select(ageg, chronic_sum) %>%
    group_by(ageg) %>% summarize(past_drinker=mean(chronic_sum))->ac2

  cerebrovascular_data %>% filter(!is.na(alcohol)) %>% filter(!is.na(chronic_sum)) %>%
    filter(alcohol==3) %>% select(ageg, chronic_sum) %>%
    group_by(ageg) %>% summarize(non_drinker=mean(chronic_sum))->ac3

  left_join(ac1,ac2, by='ageg')->drink
  left_join(drink,ac3, by='ageg')->drink

  gather(drink, key="drink", value="mean",`drinker`,`past_drinker`,`non_drinker`) %>%
  ggplot(aes(factor(ageg),mean,fill=drink))+geom_col(position="dodge")+
    scale_fill_discrete(breaks=c("drinker","past_drinker","non_drinker"))+
    labs(y="Mean of Chronic disease",x="Year",title="Chronic disease & alcohol")
```


1.00
0.75
0.50
0.25
0.00
Mean of Chronic disease
~50s
50s
60s
70s
80s~
Year
smoke
smoker
past_smoker
non_smoker


Chronic disease & alcohol
1.0
0.5
Mean of Chronic disease
drink
drinker
past_drinker
non_drinker



6.4 연령별 만성질환 상위 5개

만성질환

- hypertension:고혈압 / diabetes:당뇨병/고혈당 / cancer:암 및 악성종양 / lung_disease:폐질환
- liver_disease:간질환 / heart_disease:심장질환 / cerebrovascular_disease:뇌혈관질환
- mental_illness: 정신과적 질환 / arthritis: 관절염 또는 류마티스 / prostate_disease:전립선 질환

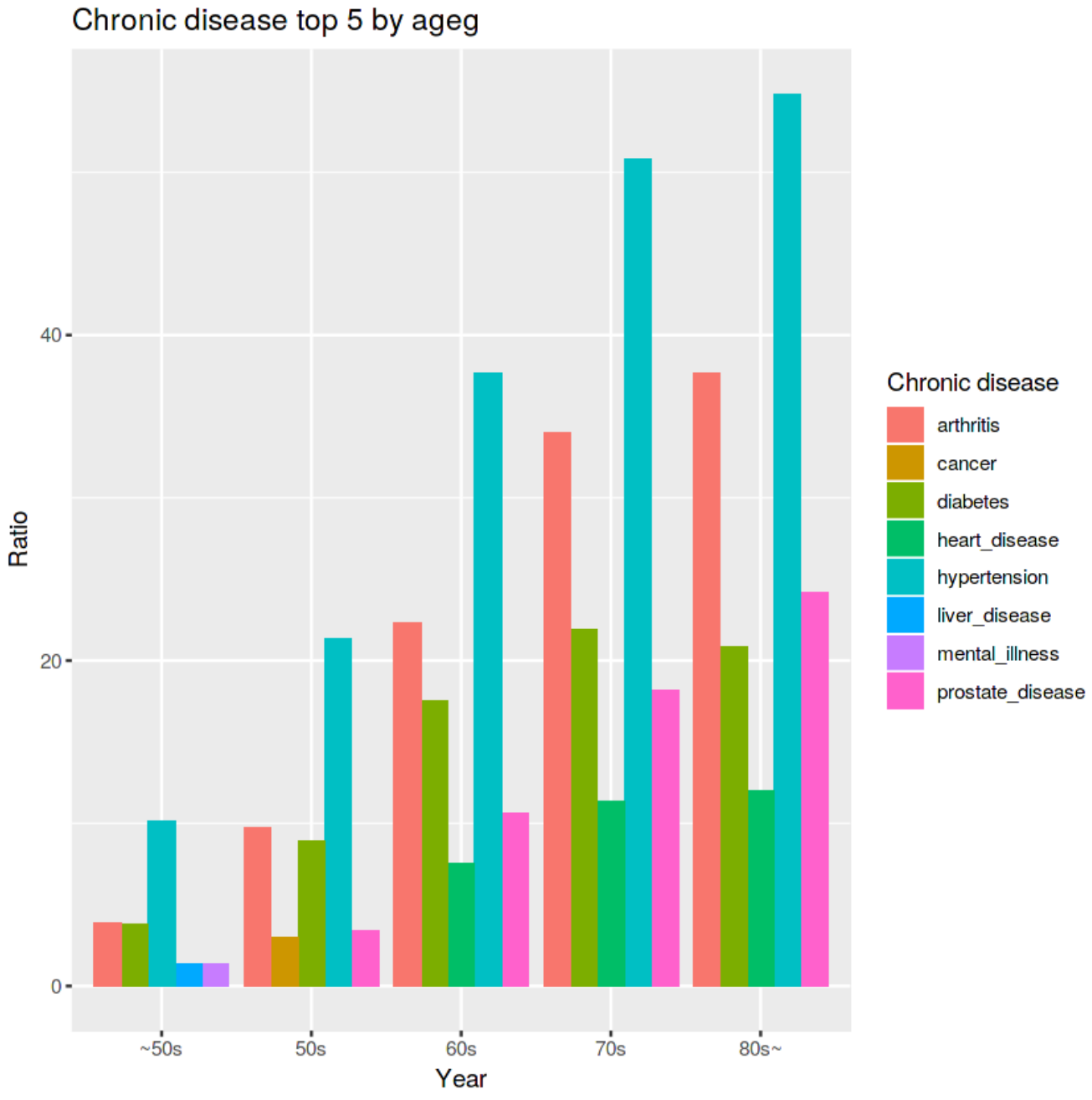
In [65]:

```
klosa  %>% filter(!is.na(hypertension))  %>% select(ageg,hypertension)  %>%
     mutate(yes=ifelse(hypertension==1,1,NA)) %>% mutate(no=ifelse(hypertension==5,1,NA)) %>%
    mutate(class="hypertension") %>% select(ageg,yes,no,class)-> t1

klosa  %>% filter(!is.na(diabetes))  %>% select(ageg,diabetes) %>%
    mutate(yes=ifelse(diabetes==1,1,NA)) %>% mutate(no=ifelse(diabetes==5,1,NA)) %>%
    mutate(class="diabetes")%>% select(ageg,yes,no,class)-> t2

klosa  %>% filter(!is.na(cancer))  %>% select(ageg,cancer) %>%
    mutate(yes=ifelse(cancer==1,1,NA)) %>% mutate(no=ifelse(cancer==5,1,NA)) %>%
    mutate(class="cancer")%>% select(ageg,yes,no,class)-> t3

klosa  %>% filter(!is.na(lung_disease))  %>% select(ageg,lung_disease) %>%
    mutate(yes=ifelse(lung_disease==1,1,NA)) %>% mutate(no=ifelse(lung_disease==5,1,NA)) %>%
    mutate(class="lung_disease")%>% select(ageg,yes,no,class)-> t4

klosa  %>% filter(!is.na(liver_disease))  %>% select(ageg,liver_disease) %>%
    mutate(yes=ifelse(liver_disease==1,1,NA)) %>% mutate(no=ifelse(liver_disease==5,1,NA)) %>%
    mutate(class="liver_disease")%>% select(ageg,yes,no,class)-> t5

klosa  %>% filter(!is.na(heart_disease))  %>% select(ageg,heart_disease) %>%
    mutate(yes=ifelse(heart_disease==1,1,NA)) %>% mutate(no=ifelse(heart_disease==5,1,NA)) %>%
    mutate(class="heart_disease")%>% select(ageg,yes,no,class)-> t6

klosa  %>% filter(!is.na(cerebrovascular_disease))  %>% select(ageg,cerebrovascular_disease) %>%
    mutate(yes=ifelse(cerebrovascular_disease==1,1,NA)) %>%
mutate(no=ifelse(cerebrovascular_disease==5,1,NA)) %>%
    mutate(class="cerebrovascular_disease")%>% select(ageg,yes,no,class)-> t7

klosa  %>% filter(!is.na(mental_illness))  %>% select(ageg,mental_illness) %>%
    mutate(yes=ifelse(mental_illness==1,1,NA)) %>% mutate(no=ifelse(mental_illness==5,1,NA)) %>%
    mutate(class="mental_illness")%>% select(ageg,yes,no,class)-> t8

klosa  %>% filter(!is.na(arthritis))  %>% select(ageg,arthritis) %>%
    mutate(yes=ifelse(arthritis==1,1,NA)) %>% mutate(no=ifelse(arthritis==5,1,NA)) %>%
    mutate(class="arthritis")%>% select(ageg,yes,no,class)-> t9

klosa  %>% filter(!is.na(prostate_disease))  %>% select(ageg,prostate_disease) %>%
    mutate(yes=ifelse(prostate_disease==1,1,NA)) %>% mutate(no=ifelse(prostate_disease==5,1,NA)) %>
%
    mutate(class="prostate_disease")%>% select(ageg,yes,no,class)-> t10

rbind(t1,t2,t3,t4,t5,t6,t7,t8,t9,t10) -> a_rbind

sqldf("select ageg,class,sum(yes)/(sum(yes)+sum(no))*100.1 as ratio from a_rbind where ageg='~50s'
group by ageg,class order by ageg,ratio desc limit 5") -> a1
sqldf("select ageg,class,sum(yes)/(sum(yes)+sum(no))*100.1 as ratio from a_rbind where ageg='50s'
```

```
sqldf("select ageg,class,sum(yes)/(sum(yes)+sum(no))*100.1 as ratio from a_rbind where ageg='50s'
group by ageg,class order by ageg,ratio desc limit 5") -> a2
sqldf("select ageg,class,sum(yes)/(sum(yes)+sum(no))*100.1 as ratio from a_rbind where ageg='60s'
group by ageg,class order by ageg,ratio desc limit 5") -> a3
sqldf("select ageg,class,sum(yes)/(sum(yes)+sum(no))*100.1 as ratio from a_rbind where ageg='70s'
group by ageg,class order by ageg,ratio desc limit 5") -> a4
sqldf("select ageg,class,sum(yes)/(sum(yes)+sum(no))*100.1 as ratio from a_rbind where ageg='80s'
group by ageg,class order by ageg,ratio desc limit 5") -> a5
sqldf("select ageg,class,sum(yes)/(sum(yes)+sum(no))*100.1 as ratio from a_rbind where ageg='80s~'
group by ageg,class order by ageg,ratio desc limit 5") -> a6
rbind(a1,a2,a3,a4,a5,a6) -> b_rbind
```

In [66]:

```
b_rbind  %>% ggplot(aes(ageg,ratio,fill=class)) +
    geom_col(position="dodge")+
    labs(y="Ratio",x="Year",title="Chronic disease top 5 by ageg") +
    theme_grey()+
    scale_fill_discrete(name = "Chronic disease")
```



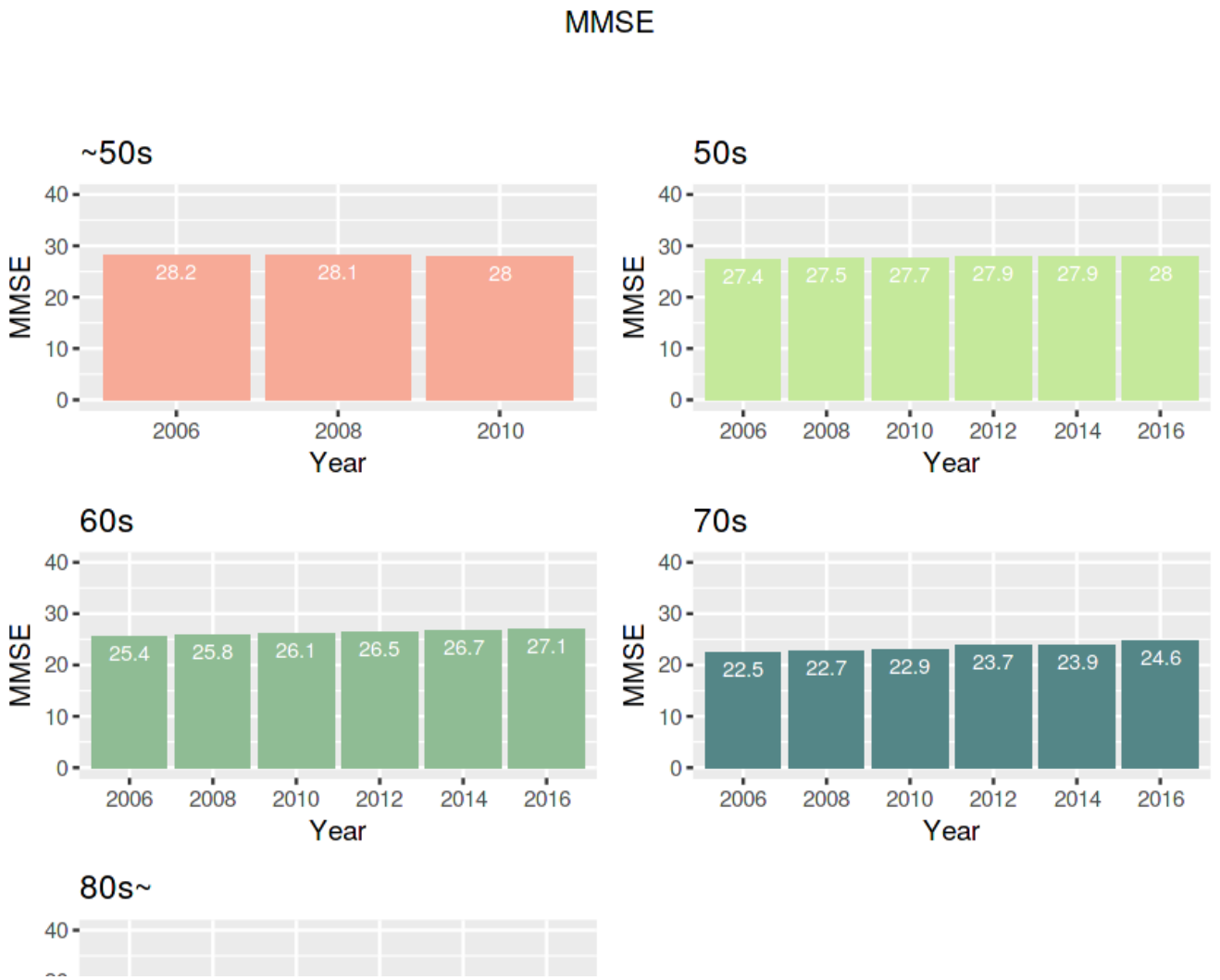
6.5 연령대별 인지기능

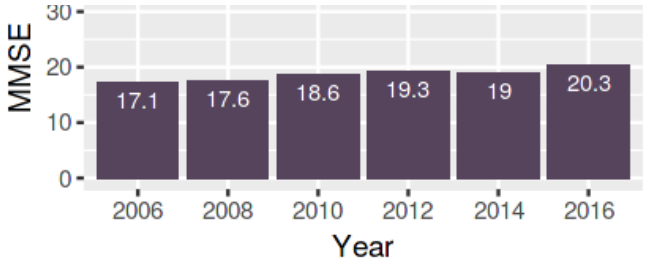
구분

- ~17 : 중증치매
- 18~23 : 의심,
- 24~ : 정상

In [67]:

```
klosa  %>% filter(!is.na(mmse))  %>% filter(ageg=='~50s') %>% select(ageg,mmse,year)  %>% group_by(
year,ageg)  %>% summarize(mean=round(mean(mmse),1)) ->k1
klosa  %>% filter(!is.na(mmse))  %>% filter(ageg=='50s') %>% select(ageg,mmse,year)  %>% group_by(y
ear,ageg)  %>% summarize(mean=round(mean(mmse),1)) ->k2
klosa  %>% filter(!is.na(mmse))  %>% filter(ageg=='60s') %>% select(ageg,mmse,year)  %>% group_by(y
ear,ageg)  %>% summarize(mean=round(mean(mmse),1)) ->k3
klosa  %>% filter(!is.na(mmse))  %>% filter(ageg=='70s') %>% select(ageg,mmse,year)  %>% group_by(y
ear,ageg)  %>% summarize(mean=round(mean(mmse),1)) ->k4
klosa  %>% filter(!is.na(mmse))  %>% filter(ageg=='80s~') %>% select(ageg,mmse,year)  %>% group_by(
year,ageg)  %>% summarize(mean=round(mean(mmse),1))->k5


ggplot(k1,aes(factor(year),mean,fill=ageg))+geom_col(position="dodge",fill="#F7AA97")+
  ylim(0,40) + geom_text(aes(label=mean), vjust=1.5, colour="white", size=3)+labs(y="MMSE",x="Year"
,title="~50s")->k1
ggplot(k2,aes(factor(year),mean,fill=ageg))+geom_col(position="dodge",fill="#C5E99B")+
  ylim(0,40) + geom_text(aes(label=mean), vjust=1.5, colour="white", size=3)+labs(y="MMSE",x="Year"
,title="50s")->k2
ggplot(k3,aes(factor(year),mean,fill=ageg))+geom_col(position="dodge",fill="#8FBC94")+
  ylim(0,40) + geom_text(aes(label=mean), vjust=1.5, colour="white", size=3)+labs(y="MMSE",x="Year"
,title="60s")->k3
ggplot(k4,aes(factor(year),mean,fill=ageg))+geom_col(position="dodge",fill="#548687")+
  ylim(0,40) + geom_text(aes(label=mean), vjust=1.5, colour="white", size=3)+labs(y="MMSE",x="Year"
,title="70s")->k4
ggplot(k5,aes(factor(year),mean,fill=ageg))+geom_col(position="dodge",fill="#56445D")+
  ylim(0,40) + geom_text(aes(label=mean), vjust=1.5, colour="white", size=3)+labs(y="MMSE",x="Year"
,title="80s~")->k5

grid.arrange(k1 ,k2 ,k3 ,k4 ,k5 , ncol=2, top="MMSE\n\n")
```



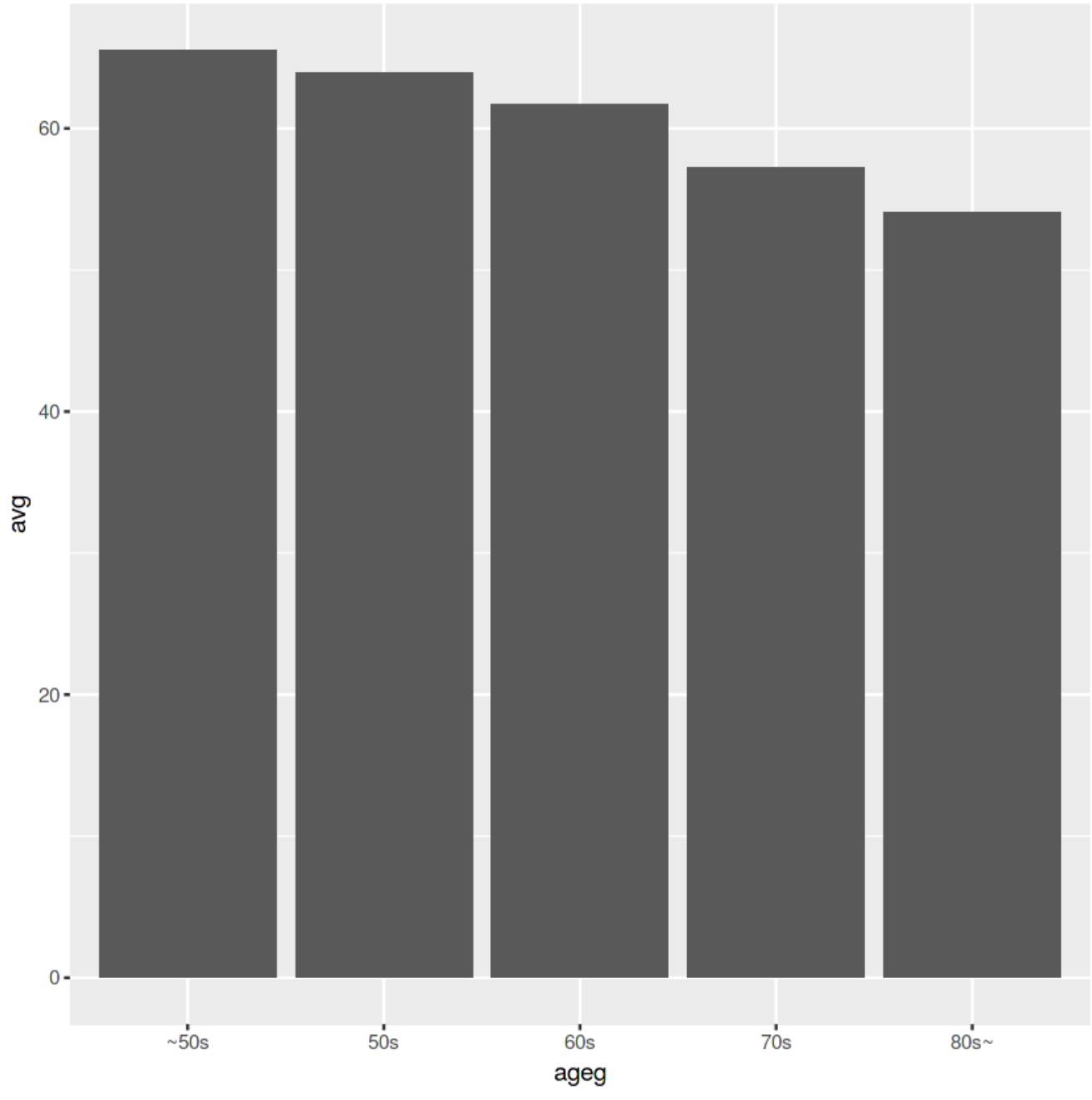


6.6 삶의 질

1) 연령대별 전반적인 삶의 질 (0 ~ 100)

In [68]:

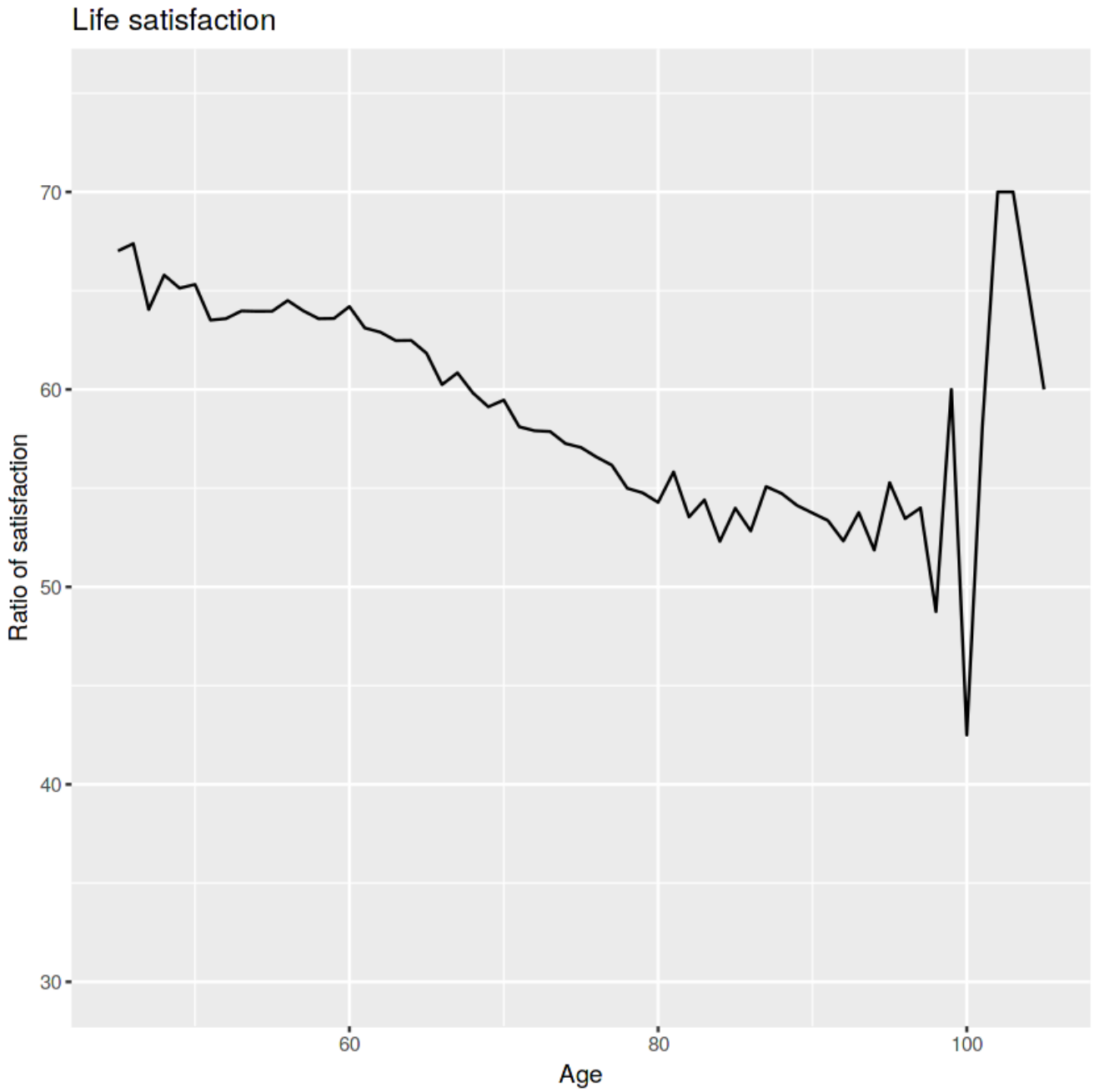
```
klosa  %>% filter(!is.na(QOL))  %>% select(ageg,QOL)-> kk1
sqldf("select ageg,avg(QOL) as avg from kk1 group by ageg") ->kk2
kk2  %>% ggplot(aes(ageg,avg)) + geom_col()
```



2) 나이별 전반적인 삶의 질 (0 ~ 100)

In [69]:

```
klosa  %>% filter(!is.na(QOL))  %>% select(age,ageg,QOL)-> satisfaction
sqldf("select age,avg(QOL) as avg from satisfaction group by age") ->satisfaction
ggplot(satisfaction,aes(age,avg))+
    geom_line()+
    labs(y="Ratio of satisfaction",x="Age",title="Life satisfaction")+
    ylim(30,75)
```



6.7 연령별, 항목별 삶의 만족도

- perceived_health: 건강상태
- perceived_eco: 경제상태
- spouse_relationship: 배우자와의 관계
- children_relationship: 자녀와의 관계

In [70]:

```
klosa %>% filter(!is.na(perceived_health)) %>%
filter(!is.na(perceived_eco)) %>%
```

```
filter(!is.na(spouse_relationship)) %>%
filter(!is.na(children_relationship)) -> mm1

library(sqldf)

sqldf("select ageg,avg(perceived_health) as a_ph, avg(perceived_eco) as a_pe,
avg(spouse_relationship) as a_sr, avg(children_relationship) as a_cr
            from mm1 group by ageg") -> pp1

pp1  %>% select(ageg,a_ph) %>% rename("value"=a_ph)  %>% mutate(class="perceived_health") -> s1
pp1  %>% select(ageg,a_pe) %>% rename("value"=a_pe)  %>% mutate(class="perceived_eco") -> s2
pp1  %>% select(ageg,a_sr) %>% rename("value"=a_sr)  %>% mutate(class="spouse_relationship") -> s3
pp1  %>% select(ageg,a_cr) %>% rename("value"=a_cr)  %>% mutate(class="children_relationship") -> s
4

rbind(s1,s2,s3,s4) -> ss1

ss1   %>%  ggplot(aes(ageg,value,fill=class)) + geom_col(position="dodge")
```