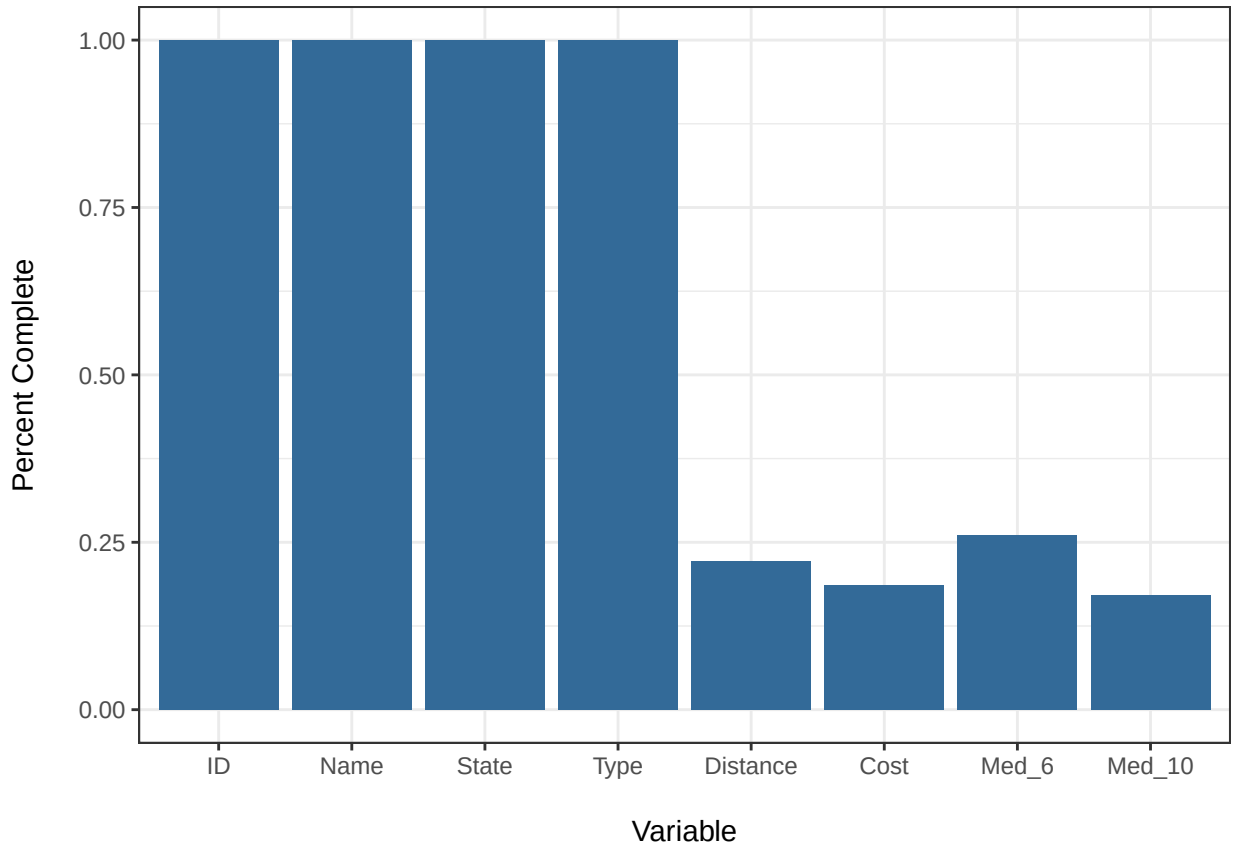
College Scorecard Data Completeness

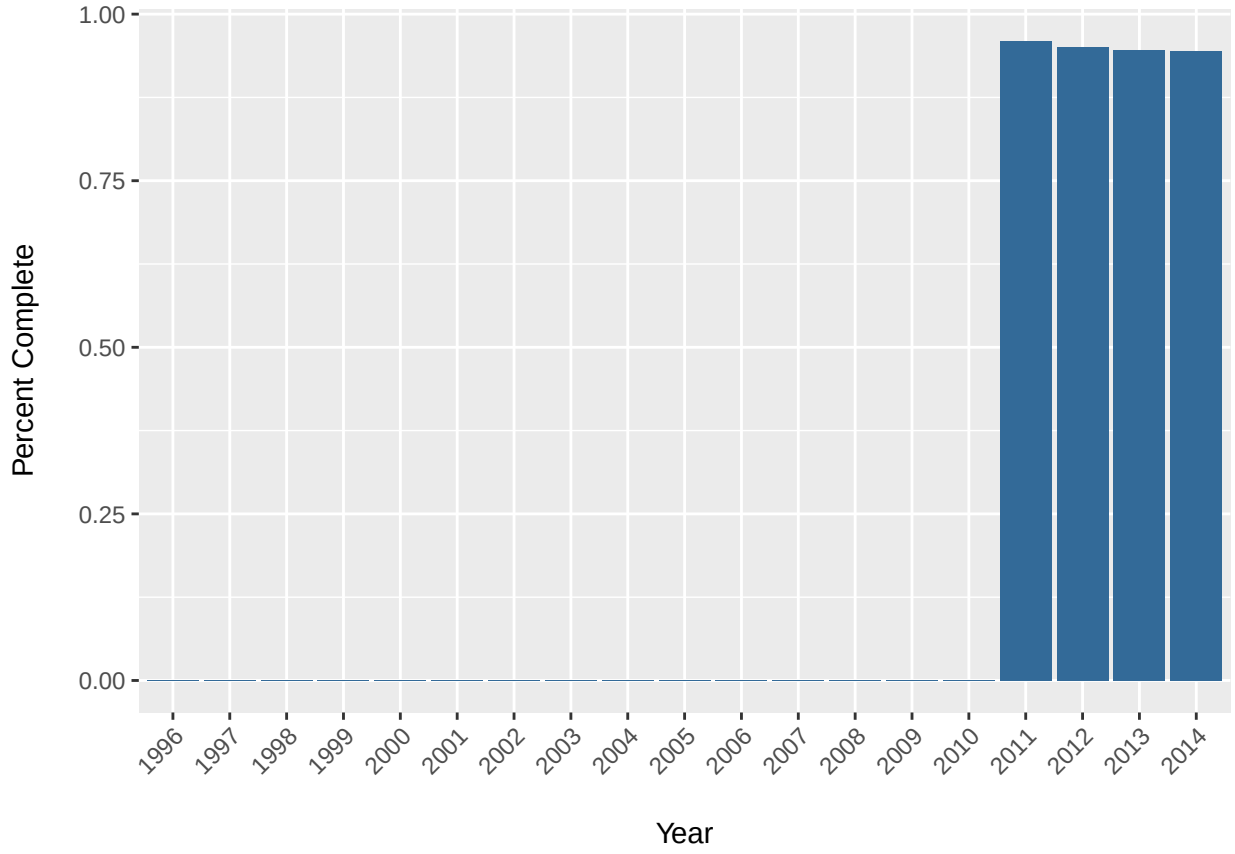
Nicole Navarro

November 3, 2016

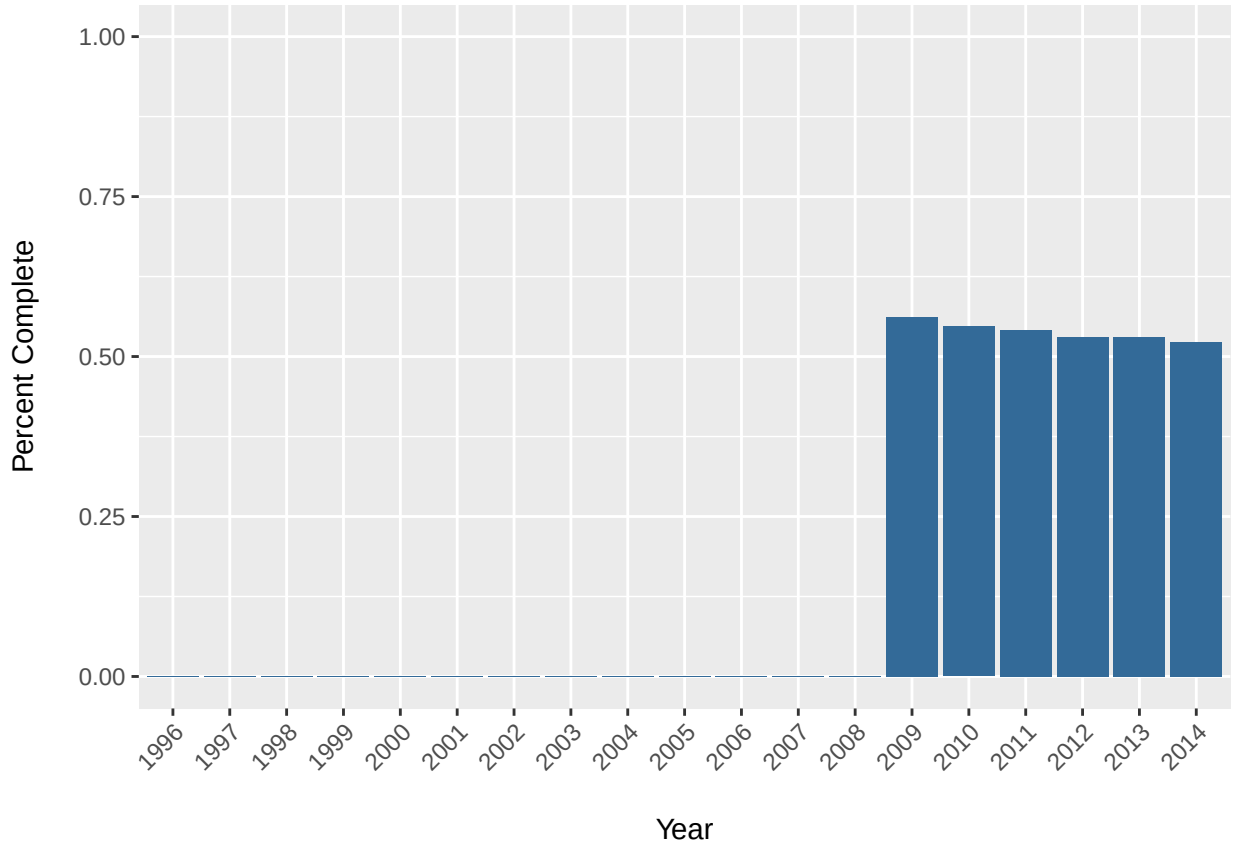
```
scorecard<-read.csv("~/School/Fall 16/EDA/scorecard_aggregated.csv")
library(ggplot2)
library(dplyr)
library(reshape2)



#calculate overall percentages of completeness

r_UNITID<-length(which(!is.na(scorecard$UNITID)))/nrow(scorecard)
r_CONTROL<-length(which(scorecard$CONTROL != "NULL"))/nrow(scorecard)
r_DISTANCEONLY<-length(which(scorecard$DISTANCEONLY != "NULL"))/nrow(scorecard)
r_INSTNM<-length(which(scorecard$INSTNM != "NULL"))/nrow(scorecard)
r_ST_FIPS<-length(which(!is.na(scorecard$ST_FIPS)))/nrow(scorecard)
r_COSTT4_A<-length(which(!is.na(scorecard$COSTT4_A)))/nrow(scorecard)
r_MD_EARN_WNE_P6<-length(which(!is.na(scorecard$MD_EARN_WNE_P6)))/nrow(scorecard)
r_MD_EARN_WNE_P10<-length(which(!is.na(scorecard$MD_EARN_WNE_P10)))/nrow(scorecard)


#Create data frame for overall completeness
ID<-c(r_UNITID)
Type<-c(r_CONTROL)
Distance<-c(r_DISTANCEONLY)
Name<-c(r_INSTNM)
State<-c(r_ST_FIPS)
Cost<-c(r_COSTT4_A)
Med_6<-c(r_MD_EARN_WNE_P6)
Med_10<-c(r_MD_EARN_WNE_P10)
pct_complete<-data.frame(ID,Name,State,Type,Distance,Cost,Med_6,Med_10)

#Create completeness plot

pct_complete$row <- seq_len(nrow(pct_complete))
pct_overall <- melt(pct_complete, id.vars = "row")


ggplot(pct_overall, aes(x=variable, y=value, fill=row)) +
  geom_bar(stat="identity") +
  xlab("\nVariable") +
  ylab("Percent Complete\n") +
  guides(fill=FALSE) +
  theme_bw()
```



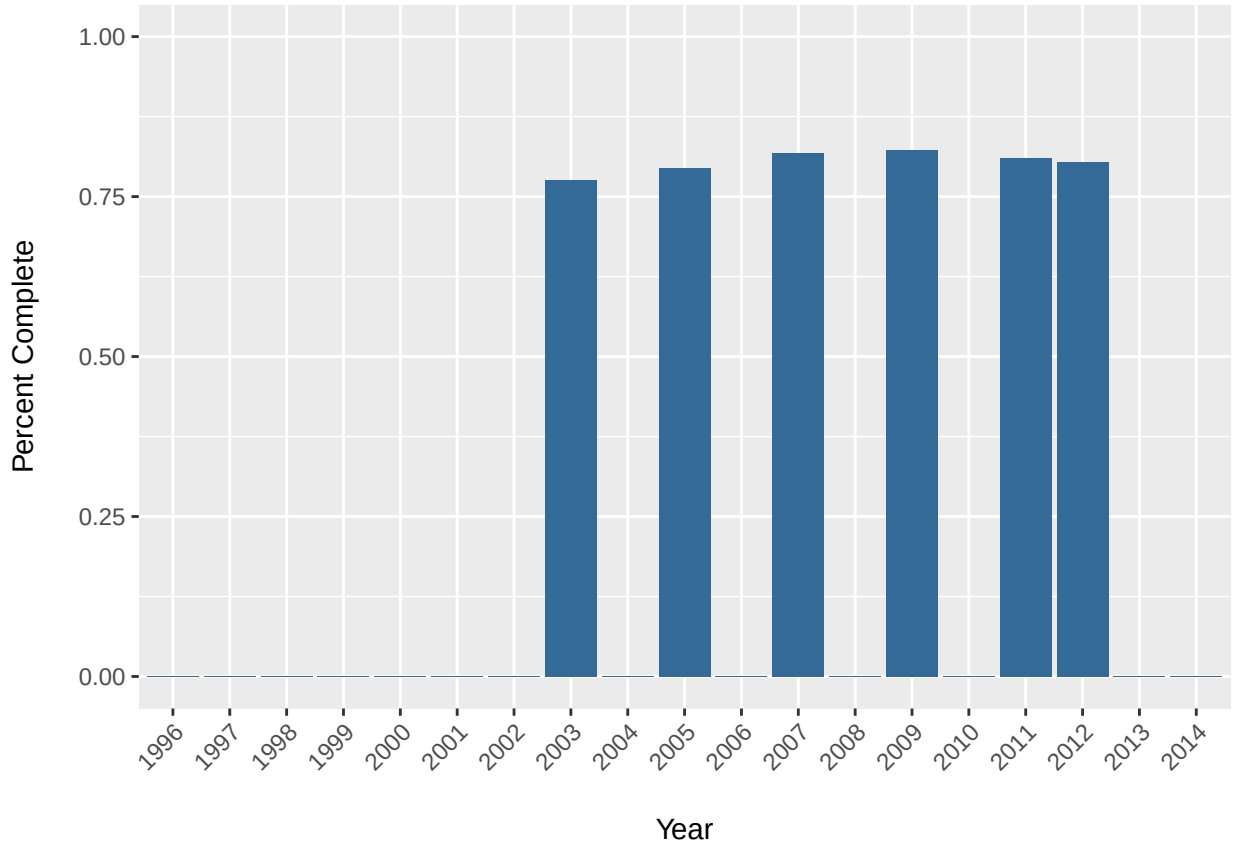

```
#completeness over years

#distance only
yr_distance<-as.data.frame(matrix(nrow=1))

for (year in 96:114){
  y<-(1900+year)
  subset<-select(scorecard,year,DISTANCEONLY)
  subset<-filter(subset,year==y)
  r<-length(which(subset$DISTANCEONLY != "NULL"))/nrow(subset)
  yr_distance[[paste(y)]]<-c(r)
}

#Create completeness plot
yr_distance<-yr_distance[-c(1)]
yr_distance$row <- seq_len(nrow(yr_distance))
pct_dist <- melt(yr_distance, id.vars = "row")

ggplot(pct_dist, aes(x=variable, y=value, fill=row)) +
  geom_bar(stat="identity") +
  xlab("\nYear") +
  ylab("Percent Complete\n") +
  guides(fill=FALSE) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



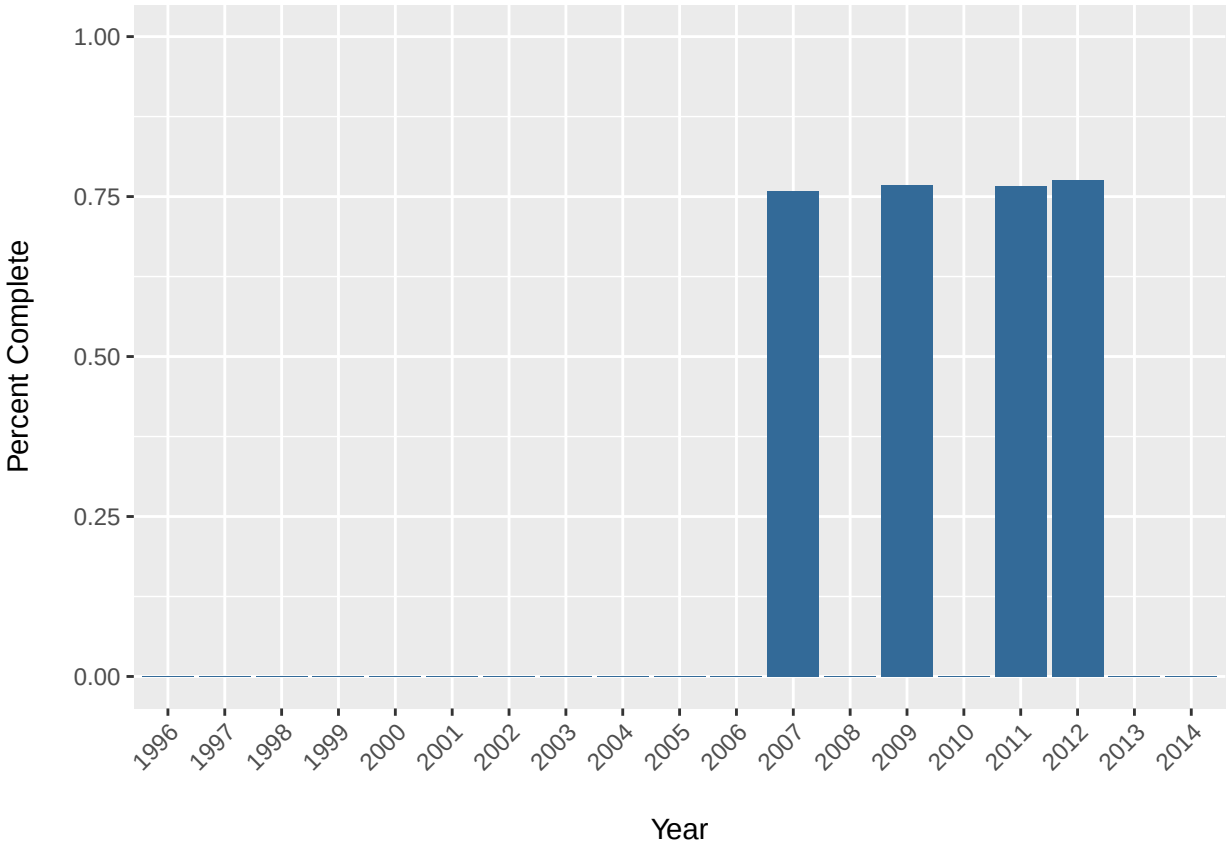

```
#Cost
yr_cost<-as.data.frame(matrix(nrow=1))

for (year in 96:114){
  y<-(1900+year)
  subset<-select(scorecard,year,COSTT4_A)
  subset<-filter(subset,year==y)
  r<-length(which(!is.na(subset$COSTT4_A)))/nrow(subset)
  yr_cost[[paste(y)]]<-c(r)
}

#Create completeness plot
yr_cost<-yr_cost[-c(1)]
yr_cost$row <- seq_len(nrow(yr_cost))
pct_cost <- melt(yr_cost, id.vars = "row")

ggplot(pct_cost, aes(x=variable, y=value, fill=row)) +
  geom_bar(stat="identity") +
  xlab("\nYear") +
  ylab("Percent Complete\n") +
  guides(fill=FALSE) +
  scale_y_continuous(limits=c(0,1)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
#Med Earnings 6
yr_med6<-as.data.frame(matrix(nrow=1))

for (year in 96:114){
  y<-(1900+year)
  subset<-select(scorecard,year,MD_EARN_WNE_P6)
  subset<-filter(subset,year==y)
  r<-length(which(!is.na(subset$MD_EARN_WNE_P6)))/nrow(subset)
  yr_med6[[paste(y)]]<-c(r)
}

#Create completeness plot
yr_med6<-yr_med6[-c(1)]
yr_med6$row <- seq_len(nrow(yr_med6))
pct_med6 <- melt(yr_med6, id.vars = "row")

ggplot(pct_med6, aes(x=variable, y=value, fill=row)) +
  geom_bar(stat="identity") +
  xlab("\nYear") +
  ylab("Percent Complete\n") +
  guides(fill=FALSE) +
  scale_y_continuous(limits=c(0,1)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
#Med Earnings 10
yr_med10<-as.data.frame(matrix(nrow=1))

for (year in 96:114){
  y<-(1900+year)
  subset<-select(scorecard,year,MD_EARN_WNE_P10)
  subset<-filter(subset,year==y)
  r<-length(which(!is.na(subset$MD_EARN_WNE_P10)))/nrow(subset)
  yr_med10[[paste(y)]]<-c(r)
}

#Create completeness plot
yr_med10<-yr_med10[-c(1)]
yr_med10$row <- seq_len(nrow(yr_med10))
pct_med10 <- melt(yr_med10, id.vars = "row")

ggplot(pct_med10, aes(x=variable, y=value, fill=row)) +
  geom_bar(stat="identity") +
  xlab("\nYear") +
  ylab("Percent Complete\n") +
  guides(fill=FALSE) +
  scale_y_continuous(limits=c(0,1)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Percent Complete
1.00
0.75
0.50
0.25
0.00
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Year