Five Refined Graphs

Ben Goldstone

9/10/2023

College Distance Data Import

College Distance Dataset

```
collegeDistance <- read_csv("~/CSVs/CollegeDistance.csv")
head(collegeDistance)
```

Composite test scores and gender

```
gf_boxplot(score~gender,data=collegeDistance, fill=~gender) + geom_violin(alpha=0.25, color=NA)
```

China Income Data Import

China Income Dataset

```
ChinaIncome <- read_csv("~/CSVs/ChinaIncome.csv")
head(ChinaIncome)
```

Construction Income over Time

```
gf_point(constructionIncome~year,data=ChinaIncome,color = ~ (constructionIncome > 1000)) +
  # Sets colors based on True or False Condition
  scale_color_manual(
    values = c("TRUE" = "cyan3", "FALSE" = "coral1"),
    # Sets labels based on condition
    labels = c("Above 1,000", "Below 1,000")) +
  # Sets legend header
  labs(color="Construction Income")
```

Industry Income and Agriculture Income

```
# Area Plot
ggplot(data=ChinaIncome,aes(x = industryIncome,y = agricultureIncome,fill="darkgreen",color="green")) +


# Sets color for industries
industry = "blue"
agriculture="darkgreen"

# Plots lines for industry and agriculture layered
gf_line(industryIncome~year,data=ChinaIncome,color=industry) %>% gf_line(agricultureIncome~year,data=Chi
```

```
# Puts "Industry" and "Agriculture" text on plot with the corresponding color
# wish I could make the font thinner or more spread out...
gf_text(x=1988,y=4750,label="industry",color=industry,hjust = 1.2) %>% gf_text(x=1988,y=500,color=agricu
```

Consumer Goods Data Import

Consumer Goods Dataset

```
ConsumerGood <- read_csv("~/CSVs/ConsumerGood.csv")
head(ConsumerGood)
```

Share and distribution of consumer goods

```
# Modify Dataset to include Ypred
ConsumerGood = ConsumerGood %>% mutate(Ypred=-0.8896+3.547*distribution)

# Linear Model
dist_share_model <- lm(share ~ distribution, data = ConsumerGood)
summary(dist_share_model)

# scatter plot
gf_point(
  share ~ distribution,
  data = ConsumerGood,
  # sets color based on residual
  color = ifelse(residuals(dist_share_model) == 0,"green",
  # if point not on line set as red or blue
  ifelse(residuals(dist_share_model) > 0, "red", "blue")
  )) %>%
  # plots line of best fit
  gf_lm(color="gray50",alpha=0.5) %>%
  # plots dotted residual lines from point to linear model
  gf_linerange(Ypred+share~distribution,linetype="dotted",color="gray30")
```