

# ETC 2420/5242 Lab 1 2017

*Di Cook*  
*SOLUTION*

## Question 1 (2pts)

There could be many possible questions that might be answered by this data. Examples include these ones:

- Does the personal savings rate dip when unemployment is high?
- Is there a seasonal effect in unemployment?
- Is population increasing?

## Question 2 (2pts)

There could be many possible questions that might be answered by this data. Examples include these ones:

- Is life expectancy positively associated with gdp percapita?
- Is life expectancy increasing over time?
- Is the trend in life expectancy similar across all countries?

## Question 3 (2pts)

There could be many possible questions that might be answered by this data. Examples include these ones:

- What places in the city see the most pedestrians?
- What times would be rush hours on week days?
- Can you see the Wednesday night markets location and time based on pedestrian traffic?
- Is White Night visible in terms of pedestrian traffic?
- Are more people out and about in summer than in winter?

## Question 4 (4pts)

1. Read in the OECD PISA data
2. (1pt) Tabulate the countries (CNT)
3. Extract the values for Australia (AUS) and Shanghai (QCN)
4. (1pt) Compute the average and standard deviation of the reading scores (PV1READ), for each country
5. (2pts) Write a few sentences explaining what you learn about reading scores in these two countries.

```
student2012.sub <- readRDS("../data/student_sub.rds")
table(student2012.sub$CNT)
#
# ARE AUS AUT BEL BGR BRA CAN CHL COL CZE DEU DNK
# 11500 14481 4755 8597 5282 5506 21544 6856 9073 5327 5001 7481
# ESP EST FIN FRA GBR HKG HRV HUN IRL ISR ITA JPN
# 10175 4779 8829 4613 4185 4670 5008 4810 5016 5055 5495 6351
# KOR MAC MNE MYS NLD NOR POL PRT QCN RUS SGP SRB
```

```

# 5033 5335 4744 5197 4460 4686 4607 5722 5177 5231 5546 4684
# SVK SVN SWE TAP TUR URY USA
# 4678 5911 4736 6046 4848 5315 4978
australia <- student2012.sub[student2012.sub$CNT=="AUS",]
shanghai <- student2012.sub[student2012.sub$CNT=="QCN",]
mean(australia$PV1READ)
# [1] 500.8453
sd(australia$PV1READ)
# [1] 100.7817
mean(shanghai$PV1READ)
# [1] 567.4197
sd(shanghai$PV1READ)
# [1] 79.91869

```

The reading scores are higher in Shanghai than in Australia by about 67 points. The variation in scores in Australia is higher, with a standard deviation of 100 as opposed to 80 for Shanghai.

```

# Alternative way to do the code
library(dplyr)
library(knitr)
library(tidyr)
student2012.sub %>% select(CNT) %>% group_by(CNT) %>% tally()
# # A tibble: 43 x 2
#   CNT      n
#   <chr> <int>
# 1  ARE 11500
# 2  AUS 14481
# 3  AUT  4755
# 4  BEL  8597
# 5  BGR  5282
# 6  BRA  5506
# 7  CAN 21544
# 8  CHL  6856
# 9  COL  9073
#10  CZE  5327
# # ... with 33 more rows
student2012.sub %>% filter(CNT %in% c("AUS", "QCN")) %>%
  group_by(CNT) %>%
  summarise(m=mean(PV1READ), s=sd(PV1READ)) %>% kable(digits=1)

```

CNT	m	s
AUS	500.8	100.8
QCN	567.4	79.9