```bash
#!/bin/bash
files="/home/wayne/pub/cs146/awkcel"

# 1(a)
    # command: use code from 1(b) to generate table, output to file named "orthologs_count.tsv"
    # then use the following commands to find most/least ortholog pair
    # least
    tail -n +2 orthologs_count.tsv | sort -k 3 -n | head -n 1
    # most
    tail -n +2 orthologs_count.tsv | sort -k 3 -nr | head -n 1

    # answer:
    # most orthologs: rat and mouse
    # least orthologs: yeast and alpaca

# 1(b)
    # command that creates a table of the number of orthologs for every pair
    species=(`echo $(head -n 1 "$files/orthologs.tsv") | tr "\t" " "`)
    num_species=${#species[@]}
    printf "first\tsecond\torthologs\n"
    for (( i=0; i<$num_species; i++));
    do
        for (( j=$((i+1)); j<$num_species; j++));
        do
            first=${species[$i]}
            second=${species[$j]}
            printf "%s\t%s\t" $first $second
            ./awkcel "$first!=\"_\" && $second!=\"_\" {count++} END{print count}" "$files/orthologs.tsv"
        done
    done
    # output: see "orthologs_count.tsv"
```

```
# 2(a)
    # command:
    ./awkcel 'date=="2011-07-01" {printf "Jul 1 2011:\nAMZN:%s\nMSFT:%s\n", AMZN, MSFT}
    date=="2011-07-05" {printf "Jul 5 2011:\nAMZN:%s\nMSFT:%s\n", AMZN, MSFT}' \
    "$files/historical.2011.tsv"

    # closing prices:
    # Jul 1 2011:
    #     AMZN:209.49
    #     MSFT:26.02
    # Jul 5 2011:
    #     AMZN:213.19
    #     MSFT:26.03
    # AMZN went up, MSFT also went up (barely)

# 2(b)
    # command
    ./awkcel '
    BEGIN{prev=0; worst=0}
    {change=AMZN-prev; prev=AMZN; if(change < worst){worst=change; day=date}}
    END{printf "worst day: %s\nchange: %.2f\n", day, worst}
    ' "$files/historical.2011.tsv"
    # answer
    # worst day: 2011-10-26
    # change: -29.07

# 2(c)
    # command
    ./awkcel '
    BEGIN{first=1;worst=0}
    {
        if (first==1) {
            first=0
            for (i=2; i<=NF; ++i) {
                prev[i] = $i  # first day, just log the prices
            }
        }
        for (i=2; i<=NF; ++i) {
            if ($i != "_" && prev[i] != "_") {
                change = $i/prev[i] * 100 - 100  # percent change
                if (change < worst) {
                    worst = change
                    day = date
                    company = i # just storing the column number of that company
                }
            }
            prev[i] = $i
        }
    }
    END{printf "worst single day loss:\ncompany(column): %s\nday: %s\nchange:
    %.2f%%\n", company, day, worst}
    ' "$files/historical.2011.tsv"
    # output:
        # worst single day loss:
        # company(column): 203
        # day: 2011-06-16
        # change: -66.70%
    # finding out which company is at column 203
    column=203
    cat "$files/historical.2011.tsv" | gawk -F"\t" "{print $column; exit}"
    # output: CSX

    # final answer:
        # worst single day loss:
        # company: CSX
        # day: 2011-06-16
```

```
# Spiral Galaxies
    # command
    ./awkcel '
    BEGIN{one=0;zero=0;half=0}
    {
        P_SP = P_CW + P_ACW
        if (P_SP == 1) {++one}
        if (P_SP == 0) {++zero}
        if (P_SP > 0.5) {++half}
    }
    END{ printf "number of galaxies that has P_SP:\nexactly 1: %d\nexactly 0: %d\n>0.5:
    %d\n", one, zero, half }
    ' "$files/SDSS+GZ1+SpArcFiRe+SFR.tsv"
    # answer
    # number of galaxies that has P_SP:
    # exactly 1: 3943
    # exactly 0: 165832
    # >0.5: 66906

    # histogram
    ./awkcel '{
    P_SP=P_CW+P_ACW;
    bin=int(10*P_SP);
    n[bin]++;
    numArcs[bin]+=totalArcLength;
    }
    END{
    for(i=0;i<=10;i++)
    print i, n[i], numArcs[i]/n[i];
    }' "$files/SDSS+GZ1+SpArcFiRe+SFR.tsv"

    # output:
    # 0 525463 192.035
    # 1 94299 329.484
    # 2 39115 449.432
    # 3 22636 510.712
    # 4 15938 544.275
    # 5 13520 565.314
    # 6 11583 589.788
    # 7 10382 614.252
    # 8 12503 640.813
    # 9 15490 691.263
    # 10 4437 730.696

```